

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
23 January 2003 (23.01.2003)

PCT

(10) International Publication Number
WO 03/006611 A2

(51) International Patent Classification⁷:

C12N

(74) Agents: **BRUNELLE, Jan et al.**, Exelixis, Inc., P.O. Box 511, 170 Harbor Way, South San Francisco, CA 94083-0511 (US).

(21) International Application Number: PCT/US02/21694

(22) International Filing Date: 10 July 2002 (10.07.2002)

(81) Designated States (*national*): AE, AG, AI, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/305,016 12 July 2001 (12.07.2001) US
60/328,507 10 October 2001 (10.10.2001) US

(71) Applicant (*for all designated States except US*): **EX-ELIXIS, INC.** [US/US]; P.O. Box 511, 170 Harbor Way, South San Francisco, CA 94083-0511 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **FRIEDMAN, Lori** [US/US]; Unit 212, One Bayside Village Place, San Francisco, CA 94107 (US). **PLOWMAN, Gregory, D.** [US/US]; 35 Winding Way, San Carlos, CA 94070 (US). **KADYK, Lisa, C.** [US/US]; 431 Mangels Avenue, San Francisco, CA 94127 (US). **LI, Danxi** [CN/US]; 90 Behr Avenue, #302, San Francisco, CA 94131 (US). **FUNKE, Roel, P.** [NL/US]; 343 California Avenue, South San Francisco, CA 94080 (US).

Published:

without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 03/006611 A2

(54) Title: GPCs AS MODIFIERS OF THE IRRTK AND p21 PATHWAYS AND METHODS OF USE

(57) Abstract: Human GPC genes are identified as modulators of the IRRTK of p21 pathways, and thus are therapeutic targets for disorders associated with defective IRRTK or p21 function. Methods for identifying modulators of IRRTK or p21, comprising screening for agents that modulate the activity of GPC are provided.

GPCs AS MODIFIERS OF THE IRRTK AND p21 PATHWAYS AND METHODS OF USE

REFERENCE TO RELATED APPLICATIONS

5 This application claims priority to U.S. provisional patent applications 60/305,016
filed 7/12/2001, and 60/328,507 filed 10/10/2001. The contents of the prior applications
are hereby incorporated in their entirety.

BACKGROUND OF THE INVENTION

10 Signal transduction pathways are made up of growth factors, their receptors, upstream
regulators of the growth factors, and downstream intracellular kinase networks. These
pathways regulate and play crucial roles in many cellular processes, such as metabolism
and proliferation.

 In humans, there are three members of the Insulin Receptor family of receptor tyrosine
15 kinases (IRRTK): Insulin Receptor (InsR), Insulin like growth factor receptor (IGFR), and
insulin receptor related receptor (IRR).

 The insulin receptor (InR) binds insulin, the major anabolic hormone in humans, and
through subsequent receptor and substrate phosphorylation, elicits a pleiotropic response,
activating multiple signaling pathways for metabolic (carbohydrate, lipid, and protein
20 processing) and growth (cell proliferation and differentiation) control (Smith RM, et al.,
Int Rev Cytol 1997;173:243-80).

 The type 1 insulin-like growth factor receptor (IGF-1R), a transmembrane tyrosine
kinase, is widely expressed across many cell types in fetal and postnatal tissues, including
motor and sensory neurons and glial cells. Activation of the receptor following binding of
25 the secreted growth factor ligands IGF-1 and IGF-2 elicits a repertoire of cellular
responses including proliferation, and the protection of cells from programmed cell death
or apoptosis. As a result, signaling through the IGF-1R is the principal pathway
responsible for somatic growth. Emerging evidence suggests that members of the IGF
family, including IGF-1, IGF-2, the IGF-1 receptor (IGF-1R), and the IGF binding
30 proteins (IGFBPs), also play important roles in the development and progression of
cancer. Both in vitro and in vivo studies show that IGFs are strong mitogens for a variety
of cancer cells. IGF-1 also has an antiapoptotic action on cancer. IGF-1R, overexpressed
in cancer cells, mediates the effects of IGFs and plays a role in cell transformation induced
by tumor virus and oncogene products (Yu H, Berkel H. J La State Med Soc 1999

Apr;151(4):218-23). Accordingly, genes identified in the IGFR growth signaling pathway are of interest as targets for proliferation, apoptosis, and neuronal survival.

Insulin receptor-related receptor (IRR) is an orphan receptor in the insulin receptor (IR) family of receptor tyrosine kinases, and is primarily localized to neural crest-derived sensory neurons during embryonic development where it might play a key role in neuronal survival. In adults, it is also expressed in pancreatic beta cells. IRR has no known ligands (Hirayama I, et al., Diabetes 1999 Jun;48(6):1237-44; Tsujimoto K, et al., Neurosci Lett. 1995 Mar 24;188(2):105-8; Reinhardt RR, et al., J Neurosci. 1994 Aug;14(8):4674-83; Reinhardt RR, et al., Endocrinology. 1993 Jul;133(1):3-10).

In the *Drosophila* there is only one member of the IRRTKs, known as InR, thus subserving the function of its three human homologues. Interestingly, in *Drosophila* the insulin receptor can drive proliferation if expressed in undifferentiated tissue.

The p21/CDKN1/WAF1/CIP1 protein (El-Deiry, W. S.; et al. Cell 75: 817-825, 1993; Harper, J. W.; et al. Cell 75: 805-816, 1993; Huppi, K et al. Oncogene 9: 3017-3020, 1994) is a cell cycle control protein that inhibits cyclin-kinase activity, is tightly regulated at the transcriptional level by p21, and mediates p21 suppression of tumor cell growth. Along with p21, p21 appears to be essential for maintaining the G2 checkpoint in human cells (Bunz, F.; Dutriaux, A.; et al. Science 282:1497-1501, 1998). Sequences of P21 are well-conserved throughout evolution, and have been identified in species as diverse as human (Genbank Identifier 13643057), *Drosophila melanogaster* (GI# 1684911), *Caenorhabditis elegans* (GI#4966283), and yeast (GI#2656016).

Glypicans (GPC) are proteins with very characteristic structures that are substituted with heparan sulfate and that are linked to the cell surface via glycosylphosphatidylinositol. The modular structure of the glypicans has been highly conserved throughout evolution. Six glypicans have been identified so far in vertebrates. Mutations in *Drosophila*, humans and mice reveal a role for these cell surface molecules in the control of cell growth, migration, and differentiation (De Cat B, and David G. Semin Cell Dev Biol 2001 Apr;12(2):117-25; Higashiyama S et al (1993) J Cell Biol. 122:933-940).

Many GPCs play important roles in various disease conditions. GPC1 regulates growth factor action in pancreatic carcinoma cells and is overexpressed in human pancreatic cancer, and its expression in this cancer may enhance tumorigenic potential in vivo (Kleeff, J., et al. (1999) Pancreas 19:281-8; Kleeff, J., et al. (1998) J Clin Invest 102:1662-73; WO200023109). Levels of GPC1 are also increased in breast cancer

(WO200023109). GPC3 is expressed in hepatocarcinoma and hepatoma, and is downregulated in mesothelioma primary tumor (Hsu, H. C., et al. (1997) Cancer Res 57:5179-84; Murthy, S. S., et al. (2000) Oncogene 19:410-6). Mutations in GPC3 are associated with Simpson-Golabi-Behmel syndrome (Pilia, G., et al. (1996) Nat Genet 12, 241-7). GPC4 is expressed in various cell lines including mammary gland and breast tumor cell lines, and variant forms of GPC4 coexisting with GPC3 variant forms may explain phenotypic variations in individuals displaying Simpson-Golabi-Behmel syndrome (Veugelers, M., et al. (1998) Genomics 53:1-11). GPC4 may be involved in diseases of abnormal cell growth and behavior (WO99/37764).

The ability to manipulate the genomes of model organisms such as *Drosophila* provides a powerful means to analyze biochemical processes that, due to significant evolutionary conservation, have direct relevance to more complex vertebrate organisms. Due to a high level of gene and pathway conservation, the strong similarity of cellular processes, and the functional conservation of genes between these model organisms and mammals, identification of the involvement of novel genes in particular pathways and their functions in such model organisms can directly contribute to the understanding of the correlative pathways and methods of modulating them in mammals (see, for example, Mechler BM et al., 1985 EMBO J 4:1551-1557; Gateff E. 1982 Adv. Cancer Res. 37: 33-74; Watson KL, et al., 1994 J Cell Sci. 18: 19-33; Miklos GL, and Rubin GM. 1996 Cell 86:521-529; Wassarman DA, et al., 1995 Curr Opin Gen Dev 5: 44-50; and Booth DR. 1999 Cancer Metastasis Rev. 18: 261-284). For example, a genetic screen can be carried out in an invertebrate model organism having underexpression (e.g. knockout) or overexpression of a gene (referred to as a "genetic entry point") that yields a visible phenotype. Additional genes are mutated in a random or targeted manner. When a gene mutation changes the original phenotype caused by the mutation in the genetic entry point, the gene is identified as a "modifier" involved in the same or overlapping pathway as the genetic entry point. When the genetic entry point is an ortholog of a human gene implicated in a disease pathway, such as IIRTK or p21, modifier genes can be identified that may be attractive candidate targets for novel therapeutics.

All references cited herein, including sequence information in referenced Genbank identifier numbers and website references, are incorporated herein in their entireties.

SUMMARY OF THE INVENTION

We have discovered genes that modify the IRRTK and p21 pathways in and *Drosophila*, and identified their human orthologs, hereinafter referred to as GPC. The invention provides methods for utilizing these IRRTK or p21 modifier genes and
5 polypeptides to identify GPC-modulating agents that are candidate therapeutic agents that can be used in the treatment of disorders associated with defective or impaired IRRTK or p21 function and/or GPC function. Preferred GPC-modulating agents specifically bind to GPC polypeptides and restore IRRTK or p21 function. Other preferred GPC-modulating agents are nucleic acid modulators such as antisense oligomers and RNAi that repress
10 GPC gene expression or product activity by, for example, binding to and inhibiting the respective nucleic acid (i.e. DNA or mRNA).

GPC modulating agents may be evaluated by any convenient *in vitro* or *in vivo* assay for molecular interaction with a GPC polypeptide or nucleic acid. In one embodiment, candidate GPC modulating agents are tested with an assay system comprising a GPC
15 polypeptide or nucleic acid. Agents that produce a change in the activity of the assay system relative to controls are identified as candidate IRRTK or p21 modulating agents. The assay system may be cell-based or cell-free. GPC-modulating agents include GPC related proteins (e.g. dominant negative mutants, and biotherapeutics); GPC-specific antibodies; GPC-specific antisense oligomers and other nucleic acid modulators; and
20 chemical agents that specifically bind to or interact with GPC or compete with GPC binding partner (e.g. by binding to a GPC binding partner). In one specific embodiment, a small molecule modulator is identified using a binding assay. In specific embodiments, the screening assay system is selected from an apoptosis assay, a cell proliferation assay, an angiogenesis assay, and a hypoxic induction assay.

25 In another embodiment, candidate IRRTK or p21 pathways modulating agents are further tested using a second assay system that detects changes in the IRRTK or p21 pathways, such as angiogenic, apoptotic, or cell proliferation changes produced by the originally identified candidate agent or an agent derived from the original agent. The second assay system may use cultured cells or non-human animals. In specific
30 embodiments, the secondary assay system uses non-human animals, including animals predetermined to have a disease or disorder implicating the IRRTK or p21 pathways, such as an angiogenic, apoptotic, or cell proliferation disorder (e.g. cancer).

The invention further provides methods for modulating the GPC function and/or the IRRTK or p21 pathways in a mammalian cell by contacting the mammalian cell with an

agent that specifically binds a GPC polypeptide or nucleic acid. The agent may be a small molecule modulator, a nucleic acid modulator, or an antibody and may be administered to a mammalian animal predetermined to have a pathology associated the IRRTK or p21 pathways.

5

DETAILED DESCRIPTION OF THE INVENTION

Genetic screens were designed to identify modifiers of the IRRTK or p21 pathways in *Drosophila*. For IRRTK, the dominant negative form of InR was expressed in the eye, resulting in a small eye phenotype. For p21, human p21 gene was overexpressed in the eye, resulting in a small, rough eye phenotype. Both screens aimed to identify enhancers or suppressors of the eye phenotype. The Dally gene was identified as a modifier of both IRRTK and p21 pathways. Accordingly, vertebrate orthologs of these modifiers, and preferably the human orthologs, glypican (GPC) genes (i.e., nucleic acids and polypeptides) are attractive drug targets for the treatment of pathologies associated with a defective IRRTK or p21 signaling pathway, such as cancer.

In vitro and in vivo methods of assessing GPC function are provided herein. Modulation of the GPC or their respective binding partners is useful for understanding the association of the IRRTK or p21 pathways and its members in normal and disease conditions and for developing diagnostics and therapeutic modalities for IRRTK or p21 related pathologies. GPC-modulating agents that act by inhibiting or enhancing GPC expression, directly or indirectly, for example, by affecting a GPC function such as binding activity, can be identified using methods provided herein. GPC modulating agents are useful in diagnosis, therapy and pharmaceutical development.

Nucleic acids and polypeptides of the invention

Sequences related to GPC nucleic acids and polypeptides that can be used in the invention are disclosed in Genbank (referenced by Genbank identifier (GI) number) as GI#s 4504080 (SEQ ID NO:1), 18567116 (SEQ ID NO:3), 13632290 (SEQ ID NO:4), 13632295 (SEQ ID NO:5), 4504082 (SEQ ID NO:6), 5360214 (SEQ ID NO:8), 3015541 (SEQ ID NO:9), 4877642 (SEQ ID NO:10), and 8051601 (SEQ ID NO:11) for nucleic acid, and GI#s 4504081 (SEQ ID NO:13), 1708021 (SEQ ID NO:14), 13632291 (SEQ ID NO:16), 4758462 (SEQ ID NO:17), 11421168 (SEQ ID NO:18), 4504083 (SEQ ID NO:19), 4758464 (SEQ ID NO:20), 9973298 (SEQ ID NO:21) and 5031719 (SEQ ID

NO:22) for polypeptides. Additionally, nucleic acid sequences of SEQ ID NOs: 2, 7, 12, 23 and 24 amino acid sequence of SEQ ID NO: 15 can be used in the invention.

GPCs are glycosylphosphatidylinositol-anchored cell surface heparan sulfate proteoglycan proteins with glypican domains. The term "GPC polypeptide" refers to a full-length GPC protein or a functionally active fragment or derivative thereof. A "functionally active" GPC fragment or derivative exhibits one or more functional activities associated with a full-length, wild-type GPC protein, such as antigenic or immunogenic activity, ability to bind natural cellular substrates, etc. The functional activity of GPC proteins, derivatives and fragments can be assayed by various methods known to one skilled in the art (Current Protocols in Protein Science (1998) Coligan *et al.*, eds., John Wiley & Sons, Inc., Somerset, New Jersey) and as further discussed below. For purposes herein, functionally active fragments also include those fragments that comprise one or more structural domains of a GPC, such as a binding domain. Protein domains can be identified using the PFAM program (Bateman A., et al., Nucleic Acids Res, 1999, 27:260-2). For example, the glypican domains (PFAM 01153) of GPC from GI#s4504081, 4758462, 4504083, 4758464, and 5031719 (SEQ ID NOs:13, 17, 19, 20, and 22, respectively) are located at approximately amino acid residues 2 to 557, 4 to 578, 1 to 555, 2 to 572, and 7 to 554, respectively. Methods for obtaining GPC polypeptides are also further described below. In some embodiments, preferred fragments are functionally active, domain-containing fragments comprising at least 25 contiguous amino acids, preferably at least 50, more preferably 75, and most preferably at least 100 contiguous amino acids of any one of SEQ ID NOs:13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 (a GPC). In further preferred embodiments, the fragment comprises the entire glypican (functionally active) domain.

The term "GPC nucleic acid" refers to a DNA or RNA molecule that encodes a GPC polypeptide. Preferably, the GPC polypeptide or nucleic acid or fragment thereof is from a human, but can also be an ortholog, or derivative thereof with at least 70% sequence identity, preferably at least 80%, more preferably 85%, still more preferably 90%, and most preferably at least 95% sequence identity with GPC. Normally, orthologs in different species retain the same function, due to presence of one or more protein motifs and/or 3-dimensional structures. Orthologs are generally identified by sequence homology analysis, such as BLAST analysis, usually using protein bait sequences. Sequences are assigned as a potential ortholog if the best hit sequence from the forward BLAST result retrieves the original query sequence in the reverse BLAST (Huynen MA and Bork P,

Proc Natl Acad Sci (1998) 95:5849-5856; Huynen MA *et al.*, Genome Research (2000) 10:1204-1210). Programs for multiple sequence alignment, such as CLUSTAL (Thompson JD *et al.*, 1994, Nucleic Acids Res 22:4673-4680) may be used to highlight conserved regions and/or residues of orthologous proteins and to generate phylogenetic trees. In a phylogenetic tree representing multiple homologous sequences from diverse species (e.g., retrieved through BLAST analysis), orthologous sequences from two species generally appear closest on the tree with respect to all other sequences from these two species. Structural threading or other analysis of protein folding (e.g., using software by ProCeryon, Biosciences, Salzburg, Austria) may also identify potential orthologs. In evolution, when a gene duplication event follows speciation, a single gene in one species, such as *Drosophila*, may correspond to multiple genes (paralogs) in another, such as human. As used herein, the term "orthologs" encompasses paralogs. As used herein, "percent (%) sequence identity" with respect to a subject sequence, or a specified portion of a subject sequence, is defined as the percentage of nucleotides or amino acids in the candidate derivative sequence identical with the nucleotides or amino acids in the subject sequence (or specified portion thereof), after aligning the sequences and introducing gaps, if necessary to achieve the maximum percent sequence identity, as generated by the program WU-BLAST-2.0a19 (Altschul *et al.*, J. Mol. Biol. (1997) 215:403-410) with all the search parameters set to default values. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A % identity value is determined by the number of matching identical nucleotides or amino acids divided by the sequence length for which the percent identity is being reported. "Percent (%) amino acid sequence similarity" is determined by doing the same calculation as for determining % amino acid sequence identity, but including conservative amino acid substitutions in addition to identical amino acids in the computation.

A conservative amino acid substitution is one in which an amino acid is substituted for another amino acid having similar properties such that the folding or activity of the protein is not significantly affected. Aromatic amino acids that can be substituted for each other are phenylalanine, tryptophan, and tyrosine; interchangeable hydrophobic amino acids are leucine, isoleucine, methionine, and valine; interchangeable polar amino acids are glutamine and asparagine; interchangeable basic amino acids are arginine, lysine and

histidine; interchangeable acidic amino acids are aspartic acid and glutamic acid; and interchangeable small amino acids are alanine, serine, threonine, cysteine and glycine.

Alternatively, an alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman (Smith and Waterman, 1981, *Advances in Applied Mathematics* 2:482-489; database: European Bioinformatics Institute; Smith and Waterman, 1981, *J. of Molec.Biol.*, 147:195-197; Nicholas et al., 1998, "A Tutorial on Searching Sequence Databases and Sequence Scoring Methods" (www.psc.cdu) and references cited therein.; W.R. Pearson, 1991, *Genomics* 11:635-650). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff (Dayhoff: *Atlas of Protein Sequences and Structure*, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA), and normalized by Gribskov (Gribskov 1986 *Nucl. Acids Res.* 14(6):6745-6763). The Smith-Waterman algorithm may be employed where default parameters are used for scoring (for example, gap open penalty of 12, gap extension penalty of two). From the data generated, the "Match" value reflects "sequence identity."

Derivative nucleic acid molecules of the subject nucleic acid molecules include sequences that hybridize to the nucleic acid sequence of any of SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 23 or 24. The stringency of hybridization can be controlled by temperature, ionic strength, pH, and the presence of denaturing agents such as formamide during hybridization and washing. Conditions routinely used are set out in readily available procedure texts (*e.g.*, *Current Protocol in Molecular Biology*, Vol. 1, Chap. 2.10, John Wiley & Sons, Publishers (1994); Sambrook *et al.*, *Molecular Cloning*, Cold Spring Harbor (1989)). In some embodiments, a nucleic acid molecule of the invention is capable of hybridizing to a nucleic acid molecule containing the nucleotide sequence of any one of SEQ ID NOs:1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 under stringent hybridization conditions that comprise: prehybridization of filters containing nucleic acid for 8 hours to overnight at 65° C in a solution comprising 6X single strength citrate (SSC) (1X SSC is 0.15 M NaCl, 0.015 M Na citrate; pH 7.0), 5X Denhardt's solution, 0.05% sodium pyrophosphate and 100 µg/ml herring sperm DNA; hybridization for 18-20 hours at 65° C in a solution containing 6X SSC, 1X Denhardt's solution, 100 µg/ml yeast tRNA and 0.05% sodium pyrophosphate; and washing of filters at 65° C for 1h in a solution containing 0.2X SSC and 0.1% SDS (sodium dodecyl sulfate).

In other embodiments, moderately stringent hybridization conditions are used that comprise: pretreatment of filters containing nucleic acid for 6 h at 40° C in a solution

containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH7.5), 5mM EDTA, 0.1% PVP, 0.1% Ficoll, 1% BSA, and 500 μ g/ml denatured salmon sperm DNA; hybridization for 18-20h at 40° C in a solution containing 35% formamide, 5X SSC, 50 mM Tris-HCl (pH7.5), 5mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 μ g/ml salmon sperm DNA, and 10% (wt/vol) dextran sulfate; followed by washing twice for 1 hour at 55° C in a solution containing 2X SSC and 0.1% SDS.

Alternatively, low stringency conditions can be used that comprise: incubation for 8 hours to overnight at 37° C in a solution comprising 20% formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.6), 5X Denhardt's solution, 10% dextran sulfate, and 20 μ g/ml denatured sheared salmon sperm DNA; hybridization in the same buffer for 18 to 20 hours; and washing of filters in 1 x SSC at about 37° C for 1 hour.

Isolation, Production, Expression, and Mis-expression of GPC Nucleic Acids and Polypeptides

GPC nucleic acids and polypeptides, useful for identifying and testing agents that modulate GPC function and for other applications related to the involvement of GPC in the IRRTK or p21 pathways. GPC nucleic acids and derivatives and orthologs thereof may be obtained using any available method. For instance, techniques for isolating cDNA or genomic DNA sequences of interest by screening DNA libraries or by using polymerase chain reaction (PCR) are well known in the art. In general, the particular use for the protein will dictate the particulars of expression, production, and purification methods. For instance, production of proteins for use in screening for modulating agents may require methods that preserve specific biological activities of these proteins, whereas production of proteins for antibody generation may require structural integrity of particular epitopes. Expression of proteins to be purified for screening or antibody production may require the addition of specific tags (*e.g.*, generation of fusion proteins). Overexpression of a GPC protein for assays used to assess GPC function, such as involvement in cell cycle regulation or hypoxic response, may require expression in eukaryotic cell lines capable of these cellular activities. Techniques for the expression, production, and purification of proteins are well known in the art; any suitable means therefore may be used (*e.g.*, Higgins SJ and Hames BD (eds.) Protein Expression: A Practical Approach, Oxford University Press Inc., New York 1999; Stanbury PF et al., Principles of Fermentation Technology, 2nd edition, Elsevier Science, New York, 1995; Doonan S (ed.) Protein Purification Protocols, Humana Press, New Jersey, 1996; Coligan JE et al, Current Protocols in Protein Science

(eds.), 1999, John Wiley & Sons, New York). In particular embodiments, recombinant GPC is expressed in a cell line known to have defective p21 function such as HCT116 colon cancer cells available from American Type Culture Collection (ATCC), Manassas, VA). The recombinant cells are used in cell-based screening assay systems of the

5 invention, as described further below.

~~The nucleotide sequence encoding a GPC polypeptide can be inserted into any~~
appropriate expression vector. The necessary transcriptional and translational signals, including promoter/enhancer element, can derive from the native GPC gene and/or its flanking regions or can be heterologous. A variety of host-vector expression systems may
10 be utilized, such as mammalian cell systems infected with virus (*e.g.* vaccinia virus, adenovirus, *etc.*); insect cell systems infected with virus (*e.g.* baculovirus); microorganisms such as yeast containing yeast vectors, or bacteria transformed with bacteriophage, plasmid, or cosmid DNA. A host cell strain that modulates the expression of, modifies, and/or specifically processes the gene product may be used.

15 To detect expression of the GPC gene product, the expression vector can comprise a promoter operably linked to a GPC gene nucleic acid, one or more origins of replication, and, one or more selectable markers (*e.g.* thymidine kinase activity, resistance to antibiotics, *etc.*). Alternatively, recombinant expression vectors can be identified by assaying for the expression of the GPC gene product based on the physical or functional
20 properties of the GPC protein in *in vitro* assay systems (*e.g.* immunoassays).

The GPC protein, fragment, or derivative may be optionally expressed as a fusion, or chimeric protein product (i.e. it is joined via a peptide bond to a heterologous protein sequence of a different protein), for example to facilitate purification or detection. A chimeric product can be made by ligating the appropriate nucleic acid sequences encoding
25 the desired amino acid sequences to each other using standard methods and expressing the chimeric product. A chimeric product may also be made by protein synthetic techniques, *e.g.* by use of a peptide synthesizer (Hunkapiller *et al.*, Nature (1984) 310:105-111).

Once a recombinant cell that expresses the GPC gene sequence is identified, the gene product can be isolated and purified using standard methods (*e.g.* ion exchange, affinity,
30 and gel exclusion chromatography; centrifugation; differential solubility; electrophoresis). Alternatively, native GPC proteins can be purified from natural sources, by standard methods (*e.g.* immunoaffinity purification). Once a protein is obtained, it may be quantified and its activity measured by appropriate methods, such as immunoassay, bioassay, or other measurements of physical properties, such as crystallography.

The methods of this invention may also use cells that have been engineered for altered expression (mis-expression) of GPC or other genes associated with the IRRTK or p21 pathways. As used herein, mis-expression encompasses ectopic expression, over-expression, under-expression, and non-expression (*e.g.* by gene knock-out or blocking expression that would otherwise normally occur).

Genetically modified animals

Animal models that have been genetically modified to alter GPC expression may be used in *in vivo* assays to test for activity of a candidate IRRTK or p21 modulating agent, or to further assess the role of GPC in a IRRTK or p21 pathways process such as apoptosis or cell proliferation. Preferably, the altered GPC expression results in a detectable phenotype, such as decreased or increased levels of cell proliferation, angiogenesis, or apoptosis compared to control animals having normal GPC expression. The genetically modified animal may additionally have altered IRRTK or p21 expression (*e.g.* IRRTK or p21 knockout). Preferred genetically modified animals are mammals such as primates, rodents (preferably mice), cows, horses, goats, sheep, pigs, dogs and cats. Preferred non-mammalian species include zebrafish, *C. elegans*, and *Drosophila*. Preferred genetically modified animals are transgenic animals having a heterologous nucleic acid sequence present as an extrachromosomal element in a portion of its cells, *i.e.* mosaic animals (*sec.* for example, techniques described by Jakobovits, 1994, *Curr. Biol.* 4:761-763.) or stably integrated into its germ line DNA (*i.e.*, in the genomic sequence of most or all of its cells). Heterologous nucleic acid is introduced into the germ line of such transgenic animals by genetic manipulation of, for example, embryos or embryonic stem cells of the host animal.

Methods of making transgenic animals are well-known in the art (for transgenic mice see Brinster *et al.*, *Proc. Nat. Acad. Sci. USA* 82: 4438-4442 (1985), U.S. Pat. Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Pat. No. 4,873,191 by Wagner *et al.*, and Hogan, B., *Manipulating the Mouse Embryo*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1986); for particle bombardment see U.S. Pat. No., 4,945,050, by Sanford *et al.*; for transgenic *Drosophila* see Rubin and Spradling, *Science* (1982) 218:348-53 and U.S. Pat. No. 4,670,388; for transgenic insects see Berghammer A.J. *et al.*, *A Universal Marker for Transgenic Insects* (1999) *Nature* 402:370-371; for transgenic Zebrafish see Lin S., *Transgenic Zebrafish*, *Methods Mol Biol.* (2000);136:375-3830; for microinjection procedures for fish, amphibian eggs and birds see Houdebine and Chourrout, *Experientia* (1991) 47:897-905; for transgenic rats see Hammer *et al.*, *Cell*

(1990) 63:1099-1112; and for culturing of embryonic stem (ES) cells and the subsequent production of transgenic animals by the introduction of DNA into ES cells using methods such as electroporation, calcium phosphate/DNA precipitation and direct injection see, e.g., 'Teratocarcinomas and Embryonic Stem Cells, A Practical Approach, E. J. Robertson, 5 ed., IRL Press (1987)). Clones of the nonhuman transgenic animals can be produced according to available methods (see Wilmut, I. *et al.* (1997) Nature 385:810-813; and PCT International Publication Nos. WO 97/07668 and WO 97/07669).

In one embodiment, the transgenic animal is a "knock-out" animal having a heterozygous or homozygous alteration in the sequence of an endogenous GPC gene that 10 results in a decrease of GPC function, preferably such that GPC expression is undetectable or insignificant. Knock-out animals are typically generated by homologous recombination with a vector comprising a transgene having at least a portion of the gene to be knocked out. Typically a deletion, addition or substitution has been introduced into the transgene to functionally disrupt it. The transgene can be a human gene (e.g., from a human 15 genomic clone) but more preferably is an ortholog of the human gene derived from the transgenic host species. For example, a mouse GPC gene is used to construct a homologous recombination vector suitable for altering an endogenous GPC gene in the mouse genome. Detailed methodologies for homologous recombination in mice are available (see Capecchi, Science (1989) 244:1288-1292; Joyner *et al.*, Nature (1989) 20 338:153-156). Procedures for the production of non-rodent transgenic mammals and other animals are also available (Houdebine and Chourrout, *supra*; Pursel *et al.*, Science (1989) 244:1281-1288; Simms *et al.*, BioTechnology (1988) 6:179-183). In a preferred embodiment, knock-out animals, such as mice harboring a knockout of a specific gene, may be used to produce antibodies against the human counterpart of the gene that has been 25 knocked out (Claesson MH *et al.*, (1994) Scan J Immunol 40:257-264; Declerck PJ *et al.*, (1995) J Biol Chem. 270:8397-400).

In another embodiment, the transgenic animal is a "knock-in" animal having an alteration in its genome that results in altered expression (e.g., increased (including ectopic) or decreased expression) of the GPC gene, e.g., by introduction of additional 30 copies of GPC, or by operatively inserting a regulatory sequence that provides for altered expression of an endogenous copy of the GPC gene. Such regulatory sequences include inducible, tissue-specific, and constitutive promoters and enhancer elements. The knock-in can be homozygous or heterozygous.

Transgenic nonhuman animals can also be produced that contain selected systems allowing for regulated expression of the transgene. One example of such a system that may be produced is the cre/loxP recombinase system of bacteriophage P1 (Lakso *et al.*, PNAS (1992) 89:6232-6236; U.S. Pat. No. 4,959,317). If a cre/loxP recombinase system
5 is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase. Another example of a recombinase
10 system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) Science 251:1351-1355; U.S. Pat. No. 5,654,182). In a preferred embodiment, both Cre-LoxP and Flp-Frt are used in the same system to regulate expression of the transgene, and for sequential deletion of vector sequences in the same cell (Sun X *et al.* (2000) Nat Genet 25:83-6).

15 The genetically modified animals can be used in genetic studies to further elucidate the IRRTK or p21 pathways, as animal models of disease and disorders implicating defective IRRTK or p21 function, and for *in vivo* testing of candidate therapeutic agents, such as those identified in screens described below. The candidate therapeutic agents are administered to a genetically modified animal having altered GPC function and
20 phenotypic changes are compared with appropriate control animals such as genetically modified animals that receive placebo treatment, and/or animals with unaltered GPC expression that receive candidate therapeutic agent.

In addition to the above-described genetically modified animals having altered GPC function, animal models having defective IRRTK or p21 function (and otherwise normal
25 GPC function), can be used in the methods of the present invention. For example, a IRRTK or p21 knockout mouse can be used to assess, *in vivo*, the activity of a candidate IRRTK or p21 modulating agent identified in one of the *in vitro* assays described below. p21 knockout mice are described in the literature (Umanoff H, *et al.*, Proc Natl Acad Sci U S A 1995 Feb 28;92(5):1709-13). Preferably, the candidate IRRTK or p21 modulating
30 agent when administered to a model system with cells defective in IRRTK or p21 function, produces a detectable phenotypic change in the model system indicating that the IRRTK or p21 function is restored, i.e., the cells exhibit normal cell cycle progression.

Modulating Agents

The invention provides methods to identify agents that interact with and/or modulate the function of GPC and/or the IRRTK or p21 pathways. Modulating agents identified by the methods are also part of the invention. Such agents are useful in a variety of

5 diagnostic and therapeutic applications associated with the IRRTK or p21 pathways, as well as in further analysis of the GPC protein and its contribution to the IRRTK or p21 pathways. Accordingly, the invention also provides methods for modulating the IRRTK or p21 pathways comprising the step of specifically modulating GPC activity by administering a GPC-interacting or -modulating agent.

10 As used herein, an "GPC-modulating agent" is any agent that modulated GPC function, for example, an agent that interacts with GPC to inhibit or enhance GPC activity or otherwise affect normal GPC function. GPC function can be affected at any level, including transcription, protein expression, protein localization, and cellular or extra-cellular activity. In a preferred embodiment, the GPC - modulating agent specifically

15 modulates the function of the GPC. The phrases "specific modulating agent", "specifically modulates", etc., are used herein to refer to modulating agents that directly bind to the GPC polypeptide or nucleic acid, and preferably inhibit, enhance, or otherwise alter, the function of the GPC. These phrases also encompasses modulating agents that alter the interaction of the GPC with a binding partner, substrate, or cofactor (e.g. by

20 binding to a binding partner of a GPC, or to a protein/binding partner complex, and altering GPC function). In a further preferred embodiment, the GPC- modulating agent is a modulator of the IRRTK or p21 pathways (e.g. it restores and/or upregulates IRRTK or p21 function) and thus is also a IRRTK or p21-modulating agent.

Preferred GPC-modulating agents include small molecule compounds; GPC-

25 interacting proteins, including antibodies and other biotherapeutics; and nucleic acid modulators such as antisense and RNA inhibitors. The modulating agents may be formulated in pharmaceutical compositions, for example, as compositions that may comprise other active ingredients, as in combination therapy, and/or suitable carriers or excipients. Techniques for formulation and administration of the compounds may be

30 found in "Remington's Pharmaceutical Sciences" Mack Publishing Co., Easton, PA, 19th edition.

Small molecule modulators

Small molecules, are often preferred to modulate function of proteins with enzymatic function, and/or containing protein interaction domains. Chemical agents, referred to in the art as "small molecule" compounds are typically organic, non-peptide molecules, having a molecular weight less than 10,000, preferably less than 5,000, more preferably less than 1,000, and most preferably less than 500. This class of modulators includes chemically synthesized molecules, for instance, compounds from combinatorial chemical libraries. Synthetic compounds may be rationally designed or identified based on known or inferred properties of the GPC protein or may be identified by screening compound libraries. Alternative appropriate modulators of this class are natural products, particularly secondary metabolites from organisms such as plants or fungi, which can also be identified by screening compound libraries for GPC-modulating activity. Methods for generating and obtaining compounds are well known in the art (Schreiber SL, Science (2000) 151: 1964-1969; Radmann J and Gunther J, Science (2000) 151:1947-1948).

Small molecule modulators identified from screening assays, as described below, can be used as lead compounds from which candidate clinical compounds may be designed, optimized, and synthesized. Such clinical compounds may have utility in treating pathologies associated with the IRR TK or p21 pathways. The activity of candidate small molecule modulating agents may be improved several-fold through iterative secondary functional validation, as further described below, structure determination, and candidate modulator modification and testing. Additionally, candidate clinical compounds are generated with specific regard to clinical and pharmacological properties. For example, the reagents may be derivatized and re-screened using *in vitro* and *in vivo* assays to optimize activity and minimize toxicity for pharmaceutical development.

Protein Modulators

Specific GPC-interacting proteins are useful in a variety of diagnostic and therapeutic applications related to the IRR TK or p21 pathways and related disorders, as well as in validation assays for other GPC-modulating agents. In a preferred embodiment, GPC-interacting proteins affect normal GPC function, including transcription, protein expression, protein localization, and cellular or extra-cellular activity. In another embodiment, GPC-interacting proteins are useful in detecting and providing information

about the function of GPC proteins, as is relevant to IRRTK or p21 related disorders, such as cancer (e.g., for diagnostic means).

An GPC-interacting protein may be endogenous, i.e. one that naturally interacts genetically or biochemically with a GPC, such as a member of the GPC pathway that
5 modulates GPC expression, localization, and/or activity. GPC-modulators include
~~dominant negative forms of GPC-interacting proteins and of GPC proteins themselves.~~
Yeast two-hybrid and variant screens offer preferred methods for identifying endogenous GPC-interacting proteins (Finley, R. L. et al. (1996) in DNA Cloning-Expression Systems: A Practical Approach, eds. Glover D. & Hames B. D (Oxford University Press, Oxford,
10 England), pp. 169-203; Fashema SF et al., Gene (2000) 250:1-14; Drees BL Curr Opin Chem Biol (1999) 3:64-70; Vidal M and Legrain P Nucleic Acids Res (1999) 27:919-29; and U.S. Pat. No. 5,928,868). Mass spectrometry is an alternative preferred method for the elucidation of protein complexes (reviewed in, e.g., Pandley A and Mann M, Nature (2000) 405:837-846; Yates JR 3rd, Trends Genet (2000) 16:5-8).

15 An GPC-interacting protein may be an exogenous protein, such as a GPC-specific antibody or a T-cell antigen receptor (see, e.g., Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Laboratory; Harlow and Lane (1999) Using antibodies: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press). GPC antibodies are further discussed below.

20 In preferred embodiments, a GPC-interacting protein specifically binds a GPC protein. In alternative preferred embodiments, a GPC-modulating agent binds a GPC substrate, binding partner, or cofactor.

Antibodies

25 In another embodiment, the protein modulator is a GPC specific antibody agonist or antagonist. The antibodies have therapeutic and diagnostic utilities, and can be used in screening assays to identify GPC modulators. The antibodies can also be used in dissecting the portions of the GPC pathway responsible for various cellular responses and in the general processing and maturation of the GPC.

30 Antibodies that specifically bind GPC polypeptides can be generated using known methods. Preferably the antibody is specific to a mammalian ortholog of GPC polypeptide, and more preferably, to human GPC. Antibodies may be polyclonal, monoclonal (mAbs), humanized or chimeric antibodies, single chain antibodies, Fab fragments, F(ab')₂ fragments, fragments produced by a FAb expression library, anti-

idiotypic (anti-Id) antibodies, and epitope-binding fragments of any of the above.

Epitopes of GPC which are particularly antigenic can be selected, for example, by routine screening of GPC polypeptides for antigenicity or by applying a theoretical method for selecting antigenic regions of a protein (Hopp and Wood (1981), Proc. Natl. Acad. Sci.

- 5 U.S.A. 78:3824-28; Hopp and Wood, (1983) Mol. Immunol. 20:483-89; Sutcliffe et al., (1983) Science 219:660-66) to the amino acid sequence shown in any of SEQ ID NOs:13, 14, 15, 16, 17, 18, 19, 20, 21, or 22. Monoclonal antibodies with affinities of 10^8 M^{-1} preferably 10^9 M^{-1} to 10^{10} M^{-1} , or stronger can be made by standard procedures as described (Harlow and Lane, *supra*; Goding (1986) Monoclonal Antibodies: Principles and Practice (2d ed) Academic Press, New York; and U.S. Pat. Nos. 4,381,292; 4,451,570; 10 and 4,618,577). Antibodies may be generated against crude cell extracts of GPC or substantially purified fragments thereof. If GPC fragments are used, they preferably comprise at least 10, and more preferably, at least 20 contiguous amino acids of a GPC protein. In a particular embodiment, GPC-specific antigens and/or immunogens are 15 coupled to carrier proteins that stimulate the immune response. For example, the subject polypeptides are covalently coupled to the keyhole limpet hemocyanin (KLH) carrier, and the conjugate is emulsified in Freund's complete adjuvant, which enhances the immune response. An appropriate immune system such as a laboratory rabbit or mouse is immunized according to conventional protocols.

- 20 The presence of GPC-specific antibodies is assayed by an appropriate assay such as a solid phase enzyme-linked immunosorbent assay (ELISA) using immobilized corresponding GPC polypeptides. Other assays, such as radioimmunoassays or fluorescent assays might also be used.

- Chimeric antibodies specific to GPC polypeptides can be made that contain different 25 portions from different animal species. For instance, a human immunoglobulin constant region may be linked to a variable region of a murine mAb, such that the antibody derives its biological activity from the human antibody, and its binding specificity from the murine fragment. Chimeric antibodies are produced by splicing together genes that encode the appropriate regions from each species (Morrison et al., Proc. Natl. Acad. Sci. (1984) 81:6851-6855; Neuberger et al., Nature (1984) 312:604-608; Takeda et al., Nature 30 (1985) 31:452-454). Humanized antibodies, which are a form of chimeric antibodies, can be generated by grafting complementary-determining regions (CDRs) (Carlos, T. M., J. M. Harlan. 1994. Blood 84:2068-2101) of mouse antibodies into a background of human framework regions and constant regions by recombinant DNA technology (Riechmann

LM, et al., 1988 Nature 323: 323-327). Humanized antibodies contain ~10% murine sequences and ~90% human sequences, and thus further reduce or eliminate immunogenicity, while retaining the antibody specificities (Co MS, and Queen C. 1991 Nature 351: 501-501; Morrison SL. 1992 Ann. Rev. Immun. 10:239-265). Humanized antibodies and methods of their production are well-known in the art (U.S. Pat. Nos. 5,530,101; 5,585,089; 5,693,762; and 6,180,370).

GPC-specific single chain antibodies which are recombinant, single chain polypeptides formed by linking the heavy and light chain fragments of the Fv regions via an amino acid bridge, can be produced by methods known in the art (U.S. Pat. No. 4,946,778; Bird, Science (1988) 242:423-426; Huston et al., Proc. Natl. Acad. Sci. USA (1988) 85:5879-5883; and Ward et al., Nature (1989) 334:544-546).

Other suitable techniques for antibody production involve in vitro exposure of lymphocytes to the antigenic polypeptides or alternatively to selection of libraries of antibodies in phage or similar vectors (Huse et al., Science (1989) 246:1275-1281). As used herein, T-cell antigen receptors are included within the scope of antibody modulators (Harlow and Lane, 1988, *supra*).

The polypeptides and antibodies of the present invention may be used with or without modification. Frequently, antibodies will be labeled by joining, either covalently or non-covalently, a substance that provides for a detectable signal, or that is toxic to cells that express the targeted protein (Menard S, et al., Int J. Biol Markers (1989) 4:131-134). A wide variety of labels and conjugation techniques are known and are reported extensively in both the scientific and patent literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent moieties, fluorescent emitting lanthanide metals, chemiluminescent moieties, bioluminescent moieties, magnetic particles, and the like (U.S. Pat. Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241). Also, recombinant immunoglobulins may be produced (U.S. Pat. No. 4,816,567). Antibodies to cytoplasmic polypeptides may be delivered and reach their targets by conjugation with membrane-penetrating toxin proteins (U.S. Pat. No. 6,086,900).

When used therapeutically in a patient, the antibodies of the subject invention are typically administered parenterally, when possible at the target site, or intravenously. The therapeutically effective dose and dosage regimen is determined by clinical studies. Typically, the amount of antibody administered is in the range of about 0.1 mg/kg –to about 10 mg/kg of patient weight. For parenteral administration, the antibodies are

formulated in a unit dosage injectable form (e.g., solution, suspension, emulsion) in association with a pharmaceutically acceptable vehicle. Such vehicles are inherently nontoxic and non-therapeutic. Examples are water, saline, Ringer's solution, dextrose solution, and 5% human serum albumin. Nonaqueous vehicles such as fixed oils, ethyl oleate, or liposome carriers may also be used. The vehicle may contain minor amounts of additives, such as buffers and preservatives, which enhance isotonicity and chemical stability or otherwise enhance therapeutic potential. The antibodies' concentrations in such vehicles are typically in the range of about 1 mg/ml to about 10 mg/ml. Immunotherapeutic methods are further described in the literature (US Pat. No. 5,859,206; WO0073469).

Specific biotherapeutics

In a preferred embodiment, a GPC-interacting protein may have biotherapeutic applications. Biotherapeutic agents formulated in pharmaceutically acceptable carriers and dosages may be used to activate or inhibit signal transduction pathways. This modulation may be accomplished by binding a ligand, thus inhibiting the activity of the pathway; or by binding a receptor, either to inhibit activation of, or to activate, the receptor. Alternatively, the biotherapeutic may itself be a ligand capable of activating or inhibiting a receptor. Biotherapeutic agents and methods of producing them are described in detail in U.S. Pat. No. 6,146,628.

GPC ligand(s), antibodies to the ligand(s) or the GPC itself may be used as biotherapeutics to modulate the activity of GPC in the IIRTK or p21 pathways.

Nucleic Acid Modulators

Other preferred GPC-modulating agents comprise nucleic acid molecules, such as antisense oligomers or double stranded RNA (dsRNA), which generally inhibit GPC activity. Preferred nucleic acid modulators interfere with the function of the GPC nucleic acid such as DNA replication, transcription, translocation of the GPC RNA to the site of protein translation, translation of protein from the GPC RNA, splicing of the GPC RNA to yield one or more mRNA species, or catalytic activity which may be engaged in or facilitated by the GPC RNA.

In one embodiment, the antisense oligomer is an oligonucleotide that is sufficiently complementary to a GPC mRNA to bind to and prevent translation, preferably by binding to the 5' untranslated region. GPC-specific antisense oligonucleotides, preferably range

from at least 6 to about 200 nucleotides. In some embodiments the oligonucleotide is preferably at least 10, 15, or 20 nucleotides in length. In other embodiments, the oligonucleotide is preferably less than 50, 40, or 30 nucleotides in length. The oligonucleotide can be DNA or RNA or a chimeric mixture or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotide can be modified at the base moiety, sugar moiety, or phosphate backbone. The oligonucleotide may include other appending groups such as peptides, agents that facilitate transport across the cell membrane, hybridization-triggered cleavage agents, and intercalating agents.

In another embodiment, the antisense oligomer is a phosphothioate morpholino oligomer (PMO). PMOs are assembled from four different morpholino subunits, each of which contain one of four genetic bases (A, C, G, or T) linked to a six-membered morpholine ring. Polymers of these subunits are joined by non-ionic phosphoramidate intersubunit linkages. Details of how to make and use PMOs and other antisense oligomers are well known in the art (e.g. see WO99/18193; Probst JC, Antisense Oligodeoxynucleotide and Ribozyme Design, Methods. (2000) 22(3):271-281; Summerton J, and Weller D. 1997 Antisense Nucleic Acid Drug Dev. 7:187-95; US Pat. No. 5,235,033; and US Pat No. 5,378,841).

Alternative preferred GPC nucleic acid modulators are double-stranded RNA species mediating RNA interference (RNAi). RNAi is the process of sequence-specific, post-transcriptional gene silencing in animals and plants, initiated by double-stranded RNA (dsRNA) that is homologous in sequence to the silenced gene. Methods relating to the use of RNAi to silence genes in *C. elegans*, *Drosophila*, plants, and humans are known in the art (Fire A, et al., 1998 Nature 391:806-811; Fire, A. Trends Genet. 15, 358-363 (1999); Sharp, P. A. RNA interference 2001. Genes Dev. 15, 485-490 (2001); Hammond, S. M., et al., Nature Rev. Genet. 2, 110-119 (2001); Tuschl, T. Chem. Biochem. 2, 239-245 (2001); Hamilton, A. et al., Science 286, 950-952 (1999); Hammond, S. M., et al., Nature 404, 293-296 (2000); Zamore, P. D., et al., Cell 101, 25-33 (2000); Bernstein, E., et al., Nature 409, 363-366 (2001); Elbashir, S. M., et al., Genes Dev. 15, 188-200 (2001); WO0129058; WO9932619; Elbashir SM, et al., 2001 Nature 411:494-498).

Nucleic acid modulators are commonly used as research reagents, diagnostics, and therapeutics. For example, antisense oligonucleotides, which are able to inhibit gene expression with exquisite specificity, are often used to elucidate the function of particular genes (see, for example, U.S. Pat. No. 6,165,790). Nucleic acid modulators are also used, for example, to distinguish between functions of various members of a biological pathway.

For example, antisense oligomers have been employed as therapeutic moieties in the treatment of disease states in animals and man and have been demonstrated in numerous clinical trials to be safe and effective (Milligan JF, *et al*, Current Concepts in Antisense Drug Design, J Med Chem. (1993) 36:1923-1937; Tonkinson JL *et al.*, Antisense Oligodeoxynucleotides as Clinical Therapeutic Agents, Cancer Invest. (1996) 14:54-65).
Accordingly, in one aspect of the invention, a GPC-specific nucleic acid modulator is used in an assay to further elucidate the role of the GPC in the IRRTK or p21 pathways, and/or its relationship to other members of the pathway. In another aspect of the invention, a GPC-specific antisense oligomer is used as a therapeutic agent for treatment of IRRTK or p21-related disease states.

Assay Systems

The invention provides assay systems and screening methods for identifying specific modulators of GPC activity. As used herein, an "assay system" encompasses all the components required for performing and analyzing results of an assay that detects and/or measures a particular event. In general, primary assays are used to identify or confirm a modulator's specific biochemical or molecular effect with respect to the GPC nucleic acid or protein. In general, secondary assays further assess the activity of a GPC modulating agent identified by a primary assay and may confirm that the modulating agent affects GPC in a manner relevant to the IRRTK or p21 pathways. In some cases, GPC modulators will be directly tested in a secondary assay.

In a preferred embodiment, the screening method comprises contacting a suitable assay system comprising a GPC polypeptide or nucleic acid with a candidate agent under conditions whereby, but for the presence of the agent, the system provides a reference activity (e.g. binding activity), which is based on the particular molecular event the screening method detects. A statistically significant difference between the agent-biased activity and the reference activity indicates that the candidate agent modulates GPC activity, and hence the IRRTK or p21 pathways. The GPC polypeptide or nucleic acid used in the assay may comprise any of the nucleic acids or polypeptides described above.

Primary Assays

The type of modulator tested generally determines the type of primary assay.

Primary assays for small molecule modulators

For small molecule modulators, screening assays are used to identify candidate modulators. Screening assays may be cell-based or may use a cell-free system that recreates or retains the relevant biochemical reaction of the target protein (reviewed in Sittampalam GS *et al.*, Curr Opin Chem Biol (1997) 1:384-91 and accompanying references). As used herein the term "cell-based" refers to assays using live cells, dead cells, or a particular cellular fraction, such as a membrane, endoplasmic reticulum, or mitochondrial fraction. The term "cell free" encompasses assays using substantially purified protein (either endogenous or recombinantly produced), partially purified or crude cellular extracts. Screening assays may detect a variety of molecular events, including protein-DNA interactions, protein-protein interactions (*e.g.*, receptor-ligand binding), transcriptional activity (*e.g.*, using a reporter gene), enzymatic activity (*e.g.*, via a property of the substrate), activity of second messengers, immunogenicity and changes in cellular morphology or other cellular characteristics. Appropriate screening assays may use a wide range of detection methods including fluorescent, radioactive, colorimetric, spectrophotometric, and amperometric methods, to provide a read-out for the particular molecular event detected.

Cell-based screening assays usually require systems for recombinant expression of GPC and any auxiliary proteins demanded by the particular assay. Appropriate methods for generating recombinant proteins produce sufficient quantities of proteins that retain their relevant biological activities and are of sufficient purity to optimize activity and assure assay reproducibility. Yeast two-hybrid and variant screens, and mass spectrometry provide preferred methods for determining protein-protein interactions and elucidation of protein complexes. In certain applications, when GPC-interacting proteins are used in screens to identify small molecule modulators, the binding specificity of the interacting protein to the GPC protein may be assayed by various known methods such as substrate processing (*e.g.* ability of the candidate GPC-specific binding agents to function as negative effectors in GPC-expressing cells), binding equilibrium constants (usually at least about 10^7 M^{-1} , preferably at least about 10^8 M^{-1} , more preferably at least about 10^9 M^{-1}), and immunogenicity (*e.g.* ability to elicit GPC specific antibody in a heterologous host such as a mouse, rat, goat or rabbit). For enzymes and receptors, binding may be assayed by, respectively, substrate and ligand processing.

The screening assay may measure a candidate agent's ability to specifically bind to or modulate activity of a GPC polypeptide, a fusion protein thereof, or to cells or membranes

bearing the polypeptide or fusion protein. The GPC polypeptide can be full length or a fragment thereof that retains functional GPC activity. The GPC polypeptide may be fused to another polypeptide, such as a peptide tag for detection or anchoring, or to another tag. The GPC polypeptide is preferably human GPC, or is an ortholog or derivative thereof as described above. In a preferred embodiment, the screening assay detects candidate agent-based modulation of GPC interaction with a binding target, such as an endogenous or exogenous protein or other substrate that has GPC-specific binding activity, and can be used to assess normal GPC gene function.

Suitable assay formats that may be adapted to screen for GPC modulators are known in the art. Preferred screening assays are high throughput or ultra high throughput and thus provide automated, cost-effective means of screening compound libraries for lead compounds (Fernandes PB, Curr Opin Chem Biol (1998) 2:597-603; Sundberg SA, Curr Opin Biotechnol 2000, 11:47-53). In one preferred embodiment, screening assays uses fluorescence technologies, including fluorescence polarization, time-resolved fluorescence, and fluorescence resonance energy transfer. These systems offer means to monitor protein-protein or DNA-protein interactions in which the intensity of the signal emitted from dye-labeled molecules depends upon their interactions with partner molecules (e.g., Selvin PR, Nat Struct Biol (2000) 7:730-4; Fernandes PB, *supra*; Hertzberg RP and Pope AJ, Curr Opin Chem Biol (2000) 4:445-451).

A variety of suitable assay systems may be used to identify candidate GPC and IRRTK or p21 pathways modulators (e.g. U.S. Pat. Nos. 5,550,019 and 6,133,437 (apoptosis assays), among others). Specific preferred assays are described in more detail below.

Apoptosis assays. Assays for apoptosis may be performed by terminal deoxynucleotidyl transferase-mediated digoxigenin-11-dUTP nick end labeling (TUNEL) assay. The TUNEL assay is used to measure nuclear DNA fragmentation characteristic of apoptosis (Lazebnik *et al.*, 1994, Nature 371, 346), by following the incorporation of fluorescein-dUTP (Yonehara *et al.*, 1989, J. Exp. Med. 169, 1747). Apoptosis may further be assayed by acridine orange staining of tissue culture cells (Lueas, R., et al., 1998, Blood 15:4730-41). An apoptosis assay system may comprise a cell that expresses a GPC, and that optionally has defective IRRTK or p21 function (e.g. IRRTK or p21 is over-expressed or under-expressed relative to wild-type cells). A test agent can be added to the apoptosis assay system and changes in induction of apoptosis relative to controls where no test agent is added, identify candidate IRRTK or p21 modulating agents. In some embodiments of

the invention, an apoptosis assay may be used as a secondary assay to test a candidate IRRTK or p21 modulating agents that is initially identified using a cell-free assay system. An apoptosis assay may also be used to test whether GPC function plays a direct role in apoptosis. For example, an apoptosis assay may be performed on cells that over- or under-
5 express GPC relative to wild type cells. Differences in apoptotic response compared to wild-type cells suggests that the GPC plays a direct role in the apoptotic response.
Apoptosis assays are described further in US Pat. No. 6,133,437.

Cell proliferation and cell cycle assays. Cell proliferation may be assayed via
10 bromodeoxyuridine (BRDU) incorporation. This assay identifies a cell population undergoing DNA synthesis by incorporation of BRDU into newly-synthesized DNA. Newly-synthesized DNA may then be detected using an anti-BRDU antibody (Hoshino *et al.*, 1986, Int. J. Cancer 38, 369; Campana *et al.*, 1988, J. Immunol. Meth. 107, 79), or by other means.
15 Cell Proliferation may also be examined using [³H]-thymidine incorporation (Chen, J., 1996, Oncogene 13:1395-403; Jeoung, J., 1995, J. Biol. Chem. 270:18367-73). This assay allows for quantitative characterization of S-phase DNA syntheses. In this assay, cells synthesizing DNA will incorporate [³H]-thymidine into newly synthesized DNA. Incorporation can then be measured by standard techniques such as by counting of
20 radioisotope in a scintillation counter (e.g., Beckman LS 3800 Liquid Scintillation Counter).

Cell proliferation may also be assayed by colony formation in soft agar (Sambrook *et al.*, Molecular Cloning, Cold Spring Harbor (1989)). For example, cells transformed with GPC are seeded in soft agar plates, and colonies are measured and counted after two
25 weeks incubation.

Involvement of a gene in the cell cycle may be assayed by flow cytometry (Gray JW *et al.* (1986) Int J Radiat Biol Relat Stud Phys Chem Med 49:237-55). Cells transfected with a GPC may be stained with propidium iodide and evaluated in a flow cytometer (available from Becton Dickinson).

30 Accordingly, a cell proliferation or cell cycle assay system may comprise a cell that expresses a GPC, and that optionally has defective IRRTK or p21 function (e.g. IRRTK or p21 is over-expressed or under-expressed relative to wild-type cells). A test agent can be added to the assay system and changes in cell proliferation or cell cycle relative to controls where no test agent is added, identify candidate IRRTK or p21 modulating agents. In

some embodiments of the invention, the cell proliferation or cell cycle assay may be used as a secondary assay to test a candidate IRRTK or p21 modulating agents that is initially identified using another assay system such as a cell-free assay system. A cell proliferation assay may also be used to test whether GPC function plays a direct role in cell

- 5 proliferation or cell cycle. For example, a cell proliferation or cell cycle assay may be performed on cells that over- or under-express GPC relative to wild-type cells. Differences in proliferation or cell cycle compared to wild type cells suggests that the GPC plays a direct role in cell proliferation or cell cycle.

- 10 **Angiogenesis.** Angiogenesis may be assayed using various human endothelial cell systems, such as umbilical vein, coronary artery, or dermal cells. Suitable assays include Alamar Blue based assays (available from Biosource International) to measure proliferation; migration assays using fluorescent molecules, such as the use of Becton Dickinson Falcon HTS FluoroBlock cell culture inserts to measure migration of cells
- 15 through membranes in presence or absence of angiogenesis enhancer or suppressors; and tubule formation assays based on the formation of tubular structures by endothelial cells on Matrigel® (Becton Dickinson). Accordingly, an angiogenesis assay system may comprise a cell that expresses a GPC, and that optionally has defective IRRTK or p21 function (e.g. IRRTK or p21 is over-expressed or under-expressed relative to wild-type
- 20 cells). A test agent can be added to the angiogenesis assay system and changes in angiogenesis relative to controls where no test agent is added, identify candidate IRRTK or p21 modulating agents. In some embodiments of the invention, the angiogenesis assay may be used as a secondary assay to test a candidate IRRTK or p21 modulating agents that is initially identified using another assay system. An angiogenesis assay may also be used
- 25 to test whether GPC function plays a direct role in cell proliferation. For example, an angiogenesis assay may be performed on cells that over- or under-express GPC relative to wild type cells. Differences in angiogenesis compared to wild type cells suggests that the GPC plays a direct role in angiogenesis.

- 30 **Hypoxic induction.** The alpha subunit of the transcription factor, hypoxia inducible factor-1 (HIF-1), is upregulated in tumor cells following exposure to hypoxia in vitro. Under hypoxic conditions, HIF-1 stimulates the expression of genes known to be important in tumour cell survival, such as those encoding glycolytic enzymes and VEGF. Induction of such genes by hypoxic conditions may be assayed by growing cells

transfected with GPC in hypoxic conditions (such as with 0.1% O₂, 5% CO₂, and balance N₂, generated in a Napco 7001 incubator (Precision Scientific)) and normoxic conditions, followed by assessment of gene activity or expression by Taqman®. For example, a hypoxic induction assay system may comprise a cell that expresses a GPC, and that

5 optionally has a mutated IRRTK or p21 (e.g. IRRTK or p21 is over-expressed or under-expressed relative to wild-type cells). A test agent can be added to the hypoxic induction assay system and changes in hypoxic response relative to controls where no test agent is added, identify candidate IRRTK or p21 modulating agents. In some embodiments of the invention, the hypoxic induction assay may be used as a secondary assay to test a

10 candidate IRRTK or p21 modulating agents that is initially identified using another assay system. A hypoxic induction assay may also be used to test whether GPC function plays a direct role in the hypoxic response. For example, a hypoxic induction assay may be performed on cells that over- or under-express GPC relative to wild type cells. Differences in hypoxic response compared to wild type cells suggests that the GPC plays a

15 direct role in hypoxic induction.

Cell adhesion. Cell adhesion assays measure adhesion of cells to purified adhesion proteins, or adhesion of cells to each other, in presence or absence of candidate modulating agents. Cell-protein adhesion assays measure the ability of agents to modulate

20 the adhesion of cells to purified proteins. For example, recombinant proteins are produced, diluted to 2.5g/mL in PBS, and used to coat the wells of a microtiter plate. The wells used for negative control are not coated. Coated wells are then washed, blocked with 1% BSA, and washed again. Compounds are diluted to 2× final test concentration and added to the blocked, coated wells. Cells are then added to the wells, and the unbound

25 cells are washed off. Retained cells are labeled directly on the plate by adding a membrane-permeable fluorescent dye, such as calcein-AM, and the signal is quantified in a fluorescent microplate reader.

Cell-cell adhesion assays measure the ability of agents to modulate binding of cell adhesion proteins with their native ligands. These assays use cells that naturally or

30 recombinantly express the adhesion protein of choice. In an exemplary assay, cells expressing the cell adhesion protein are plated in wells of a multiwell plate. Cells expressing the ligand are labeled with a membrane-permeable fluorescent dye, such as BCECF, and allowed to adhere to the monolayers in the presence of candidate agents.

Unbound cells are washed off, and bound cells are detected using a fluorescence plate reader.

High-throughput cell adhesion assays have also been described. In one such assay, small molecule ligands and peptides are bound to the surface of microscope slides using a microarray spotter, intact cells are then contacted with the slides, and unbound cells are washed off. In this assay, not only the binding specificity of the peptides and modulators against cell lines are determined, but also the functional cell signaling of attached cells using immunofluorescence techniques in situ on the microchip is measured (Falscy JR et al., Bioconjug Chem. 2001 May-Jun;12(3):346-53).

Primary assays for antibody modulators

For antibody modulators, appropriate primary assays test is a binding assay that tests the antibody's affinity to and specificity for the GPC protein. Methods for testing antibody affinity and specificity are well known in the art (Harlow and Lane, 1988, 1999, *supra*). The enzyme-linked immunosorbent assay (ELISA) is a preferred method for detecting GPC-specific antibodies; others include FACS assays, radioimmunoassays, and fluorescent assays.

Primary assays for nucleic acid modulators

For nucleic acid modulators, primary assays may test the ability of the nucleic acid modulator to inhibit or enhance GPC gene expression, preferably mRNA expression. In general, expression analysis comprises comparing GPC expression in like populations of cells (*e.g.*, two pools of cells that endogenously or recombinantly express GPC) in the presence and absence of the nucleic acid modulator. Methods for analyzing mRNA and protein expression are well known in the art. For instance, Northern blotting, slot blotting, ribonuclease protection, quantitative RT-PCR (*e.g.*, using the TaqMan®, PE Applied Biosystems), or microarray analysis may be used to confirm that GPC mRNA expression is reduced in cells treated with the nucleic acid modulator (*e.g.*, Current Protocols in Molecular Biology (1994) Ausubel FM *et al.*, eds., John Wiley & Sons, Inc., chapter 4; Freeman WM *et al.*, Biotechniques (1999) 26:112-125; Kallioniemi OP, Ann Med 2001, 33:142-147; Blohm DH and Guiseppi-Elie, A Curr Opin Biotechnol 2001, 12:41-47). Protein expression may also be monitored. Proteins are most commonly detected with specific antibodies or antisera directed against either the GPC protein or specific peptides.

A variety of means including Western blotting, ELISA, or in situ detection, are available (Harlow E and Lane D, 1988 and 1999, *supra*).

Secondary Assays

5 Secondary assays may be used to further assess the activity of GPC-modulating agent identified by any of the above methods to confirm that the modulating agent affects GPC in a manner relevant to the IRRTK or p21 pathways. As used herein, GPC-modulating agents encompass candidate clinical compounds or other agents derived from previously identified modulating agent. Secondary assays can also be used to test the activity of a
10 modulating agent on a particular genetic or biochemical pathway or to test the specificity of the modulating agent's interaction with GPC.

 Secondary assays generally compare like populations of cells or animals (*e.g.*, two pools of cells or animals that endogenously or recombinantly express GPC) in the presence and absence of the candidate modulator. In general, such assays test whether
15 treatment of cells or animals with a candidate GPC-modulating agent results in changes in the IRRTK or p21 pathways in comparison to untreated (or mock- or placebo-treated) cells or animals. Certain assays use "sensitized genetic backgrounds", which, as used herein, describe cells or animals engineered for altered expression of genes in the IRRTK or p21 or interacting pathways.

20

Cell-based assays

 Cell based assays may use cell lines known to have defective IRRTK or p21 function, such as HCT116 colon cancer cells with defective p21, available from American Type Culture Collection (ATCC), Manassas, VA).

25 Cell based assays may detect endogenous IRRTK or p21 pathway activity or may rely on recombinant expression of IRRTK or p21 pathway components. Any of the aforementioned assays may be used in this cell-based format. Candidate modulators are typically added to the cell media but may also be injected into cells or delivered by any other efficacious means.

30

Animal Assays

 A variety of non-human animal models of normal or defective IRRTK or p21 pathways may be used to test candidate GPC modulators. Models for defective IRRTK or p21 pathways typically use genetically modified animals that have been engineered to

mis-express (*e.g.*, over-express or lack expression in) genes involved in the IRRTK or p21 pathways. Assays generally require systemic delivery of the candidate modulators, such as by oral administration, injection, etc.

In a preferred embodiment, IRRTK or p21 pathways activity is assessed by monitoring
5 neovascularization and angiogenesis. Animal models with defective and normal IRRTK or p21 are used to test the candidate modulator's affect on GPC in Matrigel® assays. Matrigel® is an extract of basement membrane proteins, and is composed primarily of laminin, collagen IV, and heparin sulfate proteoglycan. It is provided as a sterile liquid at 4° C, but rapidly forms a solid gel at 37° C. Liquid Matrigel® is mixed with various
10 angiogenic agents, such as bFGF and VEGF, or with human tumor cells which over-express the GPC. The mixture is then injected subcutaneously(SC) into female athymic nude mice (Taconic, Germantown, NY) to support an intense vascular response. Mice with Matrigel® pellets may be dosed via oral (PO), intraperitoneal (IP), or intravenous (IV) routes with the candidate modulator. Mice are euthanized 5 - 12 days post-injection,
15 and the Matrigel® pellet is harvested for hemoglobin analysis (Sigma plasma hemoglobin kit). Hemoglobin content of the gel is found to correlate the degree of neovascularization in the gel.

In another preferred embodiment, the effect of the candidate modulator on GPC is assessed via tumorigenicity assays. In one example, xenograft human tumors are
20 implanted SC into female athymic mice, 6-7 week old, as single cell suspensions either from a pre-existing tumor or from *in vitro* culture. The tumors which express the GPC endogenously are injected in the flank, 1×10^5 to 1×10^7 cells per mouse in a volume of 100 μ L using a 27gauge needle. Mice are then ear tagged and tumors are measured twice weekly. Candidate modulator treatment is initiated on the day the mean tumor weight
25 reaches 100 mg. Candidate modulator is delivered IV, SC, IP, or PO by bolus administration. Depending upon the pharmacokinetics of each unique candidate modulator, dosing can be performed multiple times per day. The tumor weight is assessed by measuring perpendicular diameters with a caliper and calculated by multiplying the measurements of diameters in two dimensions. At the end of the experiment, the excised
30 tumors maybe utilized for biomarker identification or further analyses. For immunohistochemistry staining, xenograft tumors are fixed in 4% paraformaldehyde, 0.1M phosphate, pH 7.2, for 6 hours at 4°C, immersed in 30% sucrose in PBS, and rapidly frozen in isopentane cooled with liquid nitrogen.

Diagnostic and therapeutic uses

Specific GPC-modulating agents are useful in a variety of diagnostic and therapeutic applications where disease or disease prognosis is related to defects in the IRRTK or p21 pathways, such as angiogenic, apoptotic, or cell proliferation disorders. Accordingly, the invention also provides methods for modulating the IRRTK or p21 pathways in a cell, preferably a cell pre-determined to have defective or impaired IRRTK or p21 function (e.g. due to overexpression, underexpression, or misexpression of IRRTK or p21, or due to gene mutations), comprising the step of administering an agent to the cell that specifically modulates GPC activity. Preferably, the modulating agent produces a detectable phenotypic change in the cell indicating that the IRRTK or p21 function is restored. The phrase "function is restored", and equivalents, as used herein, means that the desired phenotype is achieved, or is brought closer to normal compared to untreated cells. For example, with restored IRRTK or p21 function, cell proliferation and/or progression through cell cycle may normalize, or be brought closer to normal relative to untreated cells. The invention also provides methods for treating disorders or disease associated with impaired IRRTK or p21 function by administering a therapeutically effective amount of a GPC -modulating agent that modulates the IRRTK or p21 pathways. The invention further provides methods for modulating GPC function in a cell, preferably a cell pre-determined to have defective or impaired GPC function, by administering a GPC -modulating agent. Additionally, the invention provides a method for treating disorders or disease associated with impaired GPC function by administering a therapeutically effective amount of a GPC -modulating agent.

The discovery that GPC is implicated in IRRTK or p21 pathways provides for a variety of methods that can be employed for the diagnostic and prognostic evaluation of diseases and disorders involving defects in the IRRTK or p21 pathways and for the identification of subjects having a predisposition to such diseases and disorders.

Various expression analysis methods can be used to diagnose whether GPC expression occurs in a particular sample, including Northern blotting, slot blotting, ribonuclease protection, quantitative RT-PCR, and microarray analysis. (*e.g.*, Current Protocols in Molecular Biology (1994) Ausubel FM *et al.*, eds., John Wiley & Sons, Inc., chapter 4; Freeman WM *et al.*, Biotechniques (1999) 26:112-125; Kallioniemi OP, Ann Med 2001, 33:142-147; Blohm and Guiseppi-Elie, Curr Opin Biotechnol 2001, 12:41-47). Tissues having a disease or disorder implicating defective IRRTK or p21 signaling that express a GPC, are identified as amenable to treatment with a GPC modulating agent. In a preferred

application, the IRRTK or p21 defective tissue overexpresses a GPC relative to normal tissue. For example, a Northern blot analysis of mRNA from tumor and normal cell lines, or from tumor and matching normal tissue samples from the same patient, using full or partial GPC cDNA sequences as probes, can determine whether particular tumors express or overexpress GPC. Alternatively, the TaqMan® is used for quantitative RT-PCR analysis of GPC expression in cell lines, normal tissues and tumor samples (PE Applied Biosystems).

Various other diagnostic methods may be performed, for example, utilizing reagents such as the GPC oligonucleotides, and antibodies directed against a GPC, as described above for: (1) the detection of the presence of GPC gene mutations, or the detection of either over- or under-expression of GPC mRNA relative to the non-disorder state; (2) the detection of either an over- or an under-abundance of GPC gene product relative to the non-disorder state; and (3) the detection of perturbations or abnormalities in the signal transduction pathway mediated by GPC.

Thus, in a specific embodiment, the invention is drawn to a method for diagnosing a disease or disorder in a patient that is associated with alterations in GPC expression, the method comprising: a) obtaining a biological sample from the patient; b) contacting the sample with a probe for GPC expression; c) comparing results from step (b) with a control; and d) determining whether step (c) indicates a likelihood of the disease or disorder. Preferably, the disease is cancer, most preferably a cancer as shown in TABLE 1. The probe may be either DNA or protein, including an antibody.

EXAMPLES

The following experimental section and examples are offered by way of illustration and not by way of limitation.

1. Drosophila screens

An EP overexpression screen was carried out in *Drosophila* to identify genes that modify the small eye phenotype resulting from expression of a dominant-negative InR in the eye. The EP collection contains a large set of *Drosophila* lines bearing P insertions that overexpress the gene in which they are inserted. Each EP line was crossed to lines expressing the dominant negative InR in the eye. Resulting progeny were examined for a change in the small eye phenotype. Sequence information surrounding the P insertion site was used to identify the overexpressed genes.

A dominant loss of function screen was carried out in *Drosophila* to identify genes that interact with the cyclin dependent kinase inhibitor, p21 (Bourne HR, et al., Nature (1990) 348(6297):125-132; Marshall CJ, Trends Genet (1991) 7(3):91-95). Expression of the p21 gene from GMR-p21 transgene (Hay, B. A., et al. (1994) Development 120:2121-2129) in the eye causes deterioration of normal eye morphology, resulting in reduced, rough eyes. Flies carrying this transgene were maintained as a stock (P-1025 F, genotype: y w; P{p21-pExp-gl-w[+]Hsp70(3'UTR)-5}). Females of this stock were crossed to a collection of males carrying piggyBac insertions (Fraser M *et al.*, Virology (1985) 145:356-361). Resulting progeny carrying both the transgene and transposons were scored for the effect of the transposon on the eye phenotype, i.e. whether the transposon enhanced or suppressed (or had no effect) the eye phenotype. All data was recorded and all modifiers were retested with a repeat of the original cross, and the retests were scored at least twice. Modifiers of the eye phenotype were identified as members of the p21 pathway.

The *Drosophila* Dally gene (Genbank Identifier number 3023638), was identified as an enhancer of the small eye phenotype in both the IRRTK and p21 screens, and hence a member of the IRRTK and p21 pathways. Human orthologs of the modifiers are referred to herein as GPC.

BLAST analysis (Altschul et al., *supra*) was employed to identify Targets from *Drosophila* modifiers. For example, representative sequences from GPC, GI#s 4504081, 4758462, 4504083, 4758464, and 5031719 (SEQ ID NOs:13, 17, 19, 20, and 22, respectively), share 24%, 24%, 22%, 22%, 28, and 23% amino acid identity, respectively, with the *Drosophila* Dally.

Various domains, signals, and functional subunits in proteins were analyzed using the PSORT (Nakai K., and Horton P., Trends Biochem Sci, 1999, 24:34-6; Kenta Nakai, Protein sorting signals and prediction of subcellular localization, Adv. Protein Chem. 54, 277-344 (2000)), PFAM (Bateman A., et al., Nucleic Acids Res, 1999, 27:260-2), SMART (Ponting CP, et al., SMART: identification and annotation of domains from signaling and extracellular protein sequences. Nucleic Acids Res. 1999 Jan 1;27(1):229-32), TM-HMM (Erik L.L. Sonnhammer, Gunnar von Heijne, and Anders Krogh: A hidden Markov model for predicting transmembrane helices in protein sequences. In Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology, p 175-182 Ed J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen Menlo Park, CA: AAAI Press, 1998), and clust (Remm M, and Sonnhammer E. Classification of transmembrane protein families in the *Caenorhabditis elegans* genome and identification of human

orthologs. Genome Res. 2000 Nov;10(11):1679-89) programs. For example, the glypican domains (PFAM 01153) of GPC from GI#s 4504081, 4758462, 4504083, 4758464, and 5031719 (SEQ ID NOs:13, 17, 19, 20, and 22, respectively) are located at approximately amino acid residues 2 to 557, 4 to 578, 1 to 555, 2 to 572, and 7 to 554, respectively

5

II. High-Throughput In-Vitro Binding Assay.

³³P-labeled GPC peptide is added in an assay buffer (100 mM KCl, 20 mM HEPES pH 7.6, 1 mM MgCl₂, 1% glycerol, 0.5% NP-40, 50 mM beta-mercaptoethanol, 1 mg/ml BSA, cocktail of protease inhibitors) along with a test agent to the wells of a Neutralite-
10 avidin coated assay plate and incubated at 25°C for 1 hour. Biotinylated substrate is then added to each well and incubated for 1 hour. Reactions are stopped by washing with PBS, and counted in a scintillation counter. Test agents that cause a difference in activity relative to control without test agent are identified as candidate IRRTK or p21 modulating agents.

15

III. Immunoprecipitations and Immunoblotting

For coprecipitation of transfected proteins, 3×10^6 appropriate recombinant cells containing the GPC proteins are plated on 10-cm dishes and transfected on the following day with expression constructs. The total amount of DNA is kept constant in each
20 transfection by adding empty vector. After 24 h, cells are collected, washed once with phosphate-buffered saline and lysed for 20 min on ice in 1 ml of lysis buffer containing 50 mM Hepes, pH 7.9, 250 mM NaCl, 20 mM -glycerophosphate, 1 mM sodium orthovanadate, 5 mM p-nitrophenyl phosphate, 2 mM dithiothreitol, protease inhibitors (complete, Roche Molecular Biochemicals), and 1% Nonidet P-40. Cellular debris is
25 removed by centrifugation twice at $15,000 \times g$ for 15 min. The cell lysate is incubated with 25 μ l of M2 beads (Sigma) for 2 h at 4 °C with gentle rocking.

After extensive washing with lysis buffer, proteins bound to the beads are solubilized by boiling in SDS sample buffer, fractionated by SDS-polyacrylamide gel electrophoresis, transferred to polyvinylidene difluoride membrane and blotted with the
30 indicated antibodies. The reactive bands are visualized with horseradish peroxidase coupled to the appropriate secondary antibodies and the enhanced chemiluminescence (ECL) Western blotting detection system (Amersham Pharmacia Biotech).

IV. Expression analysis

All cell lines used in the following experiments are NCI (National Cancer Institute) lines, and are available from ATCC (American Type Culture Collection, Manassas, VA 20110-2209). Normal and tumor tissues were obtained from Impath, UC Davis, Clontech, 5 Stratagene, and Ambion.

TaqMan analysis was used to assess expression levels of the disclosed genes in various samples.

RNA was extracted from each tissue sample using Qiagen (Valencia, CA) RNeasy kits, following manufacturer's protocols, to a final concentration of 50ng/μl. Single 10 stranded cDNA was then synthesized by reverse transcribing the RNA samples using random hexamers and 500ng of total RNA per reaction, following protocol 4304965 of Applied Biosystems (Foster City, CA).

Primers for expression analysis using TaqMan assay (Applied Biosystems, Foster City, CA) were prepared according to the TaqMan protocols, and the following criteria: a) 15 primer pairs were designed to span introns to eliminate genomic contamination, and b) each primer pair produced only one product.

Taqman reactions were carried out following manufacturer's protocols, in 25 μl total volume for 96-well plates and 10 μl total volume for 384-well plates, using 300nM primer and 250 nM probe, and approximately 25ng of cDNA. The standard curve for result 20 analysis was prepared using a universal pool of human cDNA samples, which is a mixture of cDNAs from a wide variety of tissues so that the chance that a target will be present in appreciable amounts is good. The raw data were normalized using 18S rRNA (universally expressed in all tissues and cells).

For each expression analysis, tumor tissue samples were compared with matched 25 normal tissues from the same patient. A gene was considered overexpressed in a tumor when the level of expression of the gene was 2 fold or higher in the tumor compared with its matched normal sample. In cases where normal tissue was not available, a universal pool of cDNA samples was used instead. In these cases, a gene was considered overexpressed in a tumor sample when the difference of expression levels between a 30 tumor sample and the average of all normal samples from the same tissue type was greater than 2 times the standard deviation of all normal samples (i.e., $\text{Tumor} - \text{average}(\text{all normal samples}) > 2 \times \text{STDEV}(\text{all normal samples})$).

Results are shown in Table 1. Data presented in bold indicate that greater than 50% of tested tumor samples of the tissue type indicated in row 1 exhibited over expression of the

gene listed in column 1, relative to normal samples. Underlined data indicates that between 25% to 49% of tested tumor samples exhibited over expression. A modulator identified by an assay described herein can be further validated for therapeutic effect by administration to a tumor in which the gene is overexpressed. A decrease in tumor growth confirms therapeutic utility of the modulator. Prior to treating a patient with the modulator, the likelihood that the patient will respond to treatment can be diagnosed by obtaining a tumor sample from the patient, and assaying for expression of the gene targeted by the modulator. The expression data for the gene(s) can also be used as a diagnostic marker for disease progression. The assay can be performed by expression analysis as described above, by antibody directed to the gene target, or by any other available detection method.

Table 1

	breast	.	colon	.	kidney	.	lung	.	ovary	.
GI#4504080 (SEQ ID NO:1)	0	3	<u>9</u>	<u>26</u>	4	19	<u>6</u>	<u>14</u>	0	4
GI#5665747 (SEQ ID NO:23)	0	3	6	26	<u>8</u>	<u>19</u>	3	14	2	4
GI#14763902 (SEQ ID NO:24)	0	3	<u>8</u>	<u>26</u>	2	19	1	14	0	4
GI#13632295 (SEQ ID NO:5)	<u>1</u>	<u>3</u>	<u>8</u>	<u>26</u>	4	19	1	14	<u>1</u>	<u>4</u>
GI#5360214 (SEQ ID NO:6)	<u>1</u>	<u>3</u>	0	26	1	19	0	14	3	4
GI#4877642 (SEQ ID NO:8)	<u>1</u>	<u>3</u>	<u>9</u>	<u>26</u>	<u>6</u>	<u>19</u>	3	14	0	4

WHAT IS CLAIMED IS:

1. A method of identifying a candidate IRRTK or p21 pathways modulating agent, said method comprising the steps of:
 - 5 (a) providing an assay system comprising a purified GPC polypeptide or nucleic acid or a functionally active fragment or derivative thereof;
 - (b) contacting the assay system with a test agent under conditions whereby, but for the presence of the test agent, the system provides a reference activity; and
 - (c) detecting a test agent-biased activity of the assay system, wherein a difference
10 between the test agent-biased activity and the reference activity identifies the test agent as a candidate IRRTK or p21 pathways modulating agent.
2. The method of Claim 1 wherein the assay system comprises cultured cells that express the GPC polypeptide.
- 15 3. The method of Claim 2 wherein the cultured cells additionally have defective IRRTK or p21 function.
4. The method of Claim 1 wherein the assay system includes a screening assay
20 comprising a GPC polypeptide, and the candidate test agent is a small molecule modulator.
5. The method of Claim 4 wherein the assay is a binding assay.
- 25 6. The method of Claim 1 wherein the assay system is selected from the group consisting of an apoptosis assay system, a cell proliferation assay system, an angiogenesis assay system, and a hypoxic induction assay system.
7. The method of Claim 1 wherein the assay system includes a binding assay
30 comprising a GPC polypeptide and the candidate test agent is an antibody.
8. The method of Claim 1 wherein the assay system includes an expression assay comprising a GPC nucleic acid and the candidate test agent is a nucleic acid modulator.

9. The method of claim 8 wherein the nucleic acid modulator is an antisense oligomer.

10. The method of Claim 8 wherein the nucleic acid modulator is a PMO.

5

11. The method of Claim 1 additionally comprising:

(d) administering the candidate IRRTK or p21 pathways modulating agent identified in (c) to a model system comprising cells defective in IRRTK or p21 function and, detecting a phenotypic change in the model system that indicates that the IRRTK or p21 function is restored.

10

12. The method of Claim 11 wherein the model system is a mouse model with defective IRRTK or p21 function.

13. A method for modulating a IRRTK or p21 pathways of a cell comprising contacting a cell defective in IRRTK or p21 function with a candidate modulator that specifically binds to a GPC polypeptide comprising an amino acid sequence selected from group consisting of SEQ ID NOs:13, 14, 15, 16, 17, 18, 19, 20, 21, and 22, whereby IRRTK or p21 function is restored.

15

20

14. The method of claim 13 wherein the candidate modulator is administered to a vertebrate animal predetermined to have a disease or disorder resulting from a defect in IRRTK or p21 function.

15. The method of Claim 13 wherein the candidate modulator is selected from the group consisting of an antibody and a small molecule.

25

16. The method of Claim 1, comprising the additional steps of:

(d) providing a secondary assay system comprising cultured cells or a non-human animal expressing GPC ,

30

(e) contacting the secondary assay system with the test agent of (b) or an agent derived therefrom under conditions whereby, but for the presence of the test agent or agent derived therefrom, the system provides a reference activity; and

(f) detecting an agent-biased activity of the second assay system,

wherein a difference between the agent-biased activity and the reference activity of the second assay system confirms the test agent or agent derived therefrom as a candidate IRRTK or p21 pathways modulating agent,
and wherein the second assay detects an agent-biased change in the IRRTK or p21
5 pathways.

17. The method of Claim 16 wherein the secondary assay system comprises cultured cells.

10 18. The method of Claim 16 wherein the secondary assay system comprises a non-human animal.

19. The method of Claim 18 wherein the non-human animal mis-expresses a IRRTK or p21 pathways gene.

15 20. A method of modulating IRRTK or p21 pathways in a mammalian cell comprising contacting the cell with an agent that specifically binds a GPC polypeptide or nucleic acid.

21. The method of Claim 20 wherein the agent is administered to a mammalian
20 animal predetermined to have a pathology associated with the IRRTK or p21 pathways.

22. The method of Claim 20 wherein the agent is a small molecule modulator, a nucleic acid modulator, or an antibody.

25 23. A method for diagnosing a disease in a patient comprising:
(a) obtaining a biological sample from the patient;
(b) contacting the sample with a probe for GPC expression;
(c) comparing results from step (b) with a control;
(d) determining whether step (c) indicates a likelihood of disease.

30 24. The method of claim 23 wherein said disease is cancer.

25. The method according to claim 24, wherein said cancer is a cancer as shown in Table 1 as having >25% expression level.

SEQUENCE LISTING

<110> EXELIXIS, INC.

<120> GPCs AS MODIFIERS OF THE IRRTK AND p21 PATHWAYS AND METHODS OF USE

<130> EX02-081C-PC

<150> US 60/305,016

<151> 2001-07-12

<150> US 60/328,507

<151> 2001-10-10

<160> 24

<170> PatentIn version 3.1

<210> 1

<211> 3692

<212> DNA

<213> Homo sapiens

<400> 1

```

ggctgcccga gcgagcggtc ggacctcgca ccccgcgcg ccccgcgcg cgcgcgcgc 60
ggcttttggt gtctccgct cctcgccgc cgcgcctct ggaccgagag ccgcgcgcg 120
cgggaccttg gctctgccct tcgcgggcg gaactgcga ggaccggcc aggatccgag 180
agaggcgcg gcggtggcc gggggcgcc cggcccccgc catggagctc cgggcccag 240
gctggtggct gctatgtcg gccgcagcg tggtcgctg cggccgggg gaccggcca 300
gcaagagccg gagctcggc gaggtccgc agatctacg agccaaggc ttcagcctga 360
gcgacgtgcc ccaggcgag atctcgggtg agcacctcg gatctgtccc cagggtaca 420
cctgctgcac cagcgagatg gaggagaacc tggccaacc cagccatgcc gagctggaga 480
ccgcgctccg ggacagcag cgcgtcctgc aggccatgct tgccaccag ctgcgcagct 540
tcgatgacca ctccagcac ctgctgaacg actcggagcg gacgctgcag gccacctcc 600
ccggcgctt cggagagctg tacacgcaga acgcgaggc ctccggggac ctgtactcag 660
agctgcgct gtactaccg ggtgccaacc tgcacctga ggagacgct gccgagttct 720
gggcccgcct gctcgagcg ctcttcaagc agctgcacc ccagctgctg ctgcctgatg 780
actacctga ctgcctggc aagcaggcc aggcgctcg gcccttcggg gaggccccga 840
gagagctcg cctgcgggc acccgtgct tcgtggctgc tcgtccttt gtgcagggcc 900
tgggcgctg cagcgacgtg gtccggaag tggctcaggt cccctgggc cggagtgct 960
cgagagctgt catgaagctg gtctactgt ctcactgct gggagtccc ggcgccagc 1020
cctgccctga ctattgccg aatgtgctc agggctgct tgccaaccag gccgacctgg 1080
acgccgagtg gaggaacctc ctggaactca tgggtgctcat caccgacaag ttctgggta 1140

```


catcgggtgt ggagagtgtc atcggcagcg tgcacacgtg gctggcggag gccatcaacg 1200
ccctccagga caacagggac acgtccacgg ccaaggatcat ccagggtctgc gggaacccca 1260
aggteaacc ccagggccctt gggcctgagg agaagcggcg ccggggcaag ctggccccgc 1320
gggagaggcc accttcaggc acgtctggaga agctggtctc tgaagccaag gccagctcc 1380
gcgacgtcca ggaattcttg atcagcctcc caggacact gtgcagttag aagatggccc 1440
tgagcactgc cagtgatgac cgtcgttga acgggatggc cagaggccgg tacctccccg 1500
aggatcatgg tgacggcctg gccaacaga tcaacaacc cgagggtggag gtggacatca 1560
ccaagccgga catgaccatc cggcagcaga tcatgcagct gaagatcatg accaaccggc 1620
tgcgcagcgc ctacaacggc aacgacgtgg acttcacaga cgcagtgac gacggcagcg 1680
gctcgggcag cggatgatggc tgtctggatg acctctcgg ccggaaggtc agcaggaaga 1740
gtccagctc ccggacgcc ttgacctg ccctcccagg cctgtcagag caggaaggac 1800
agaagacctc ggctgccagc tgccccagc ccccgacctt cctcctgcc ctcctcctct 1860
tcctggccct tacagtagcc agggcccggg ggcggtaact gcccgaagg cccagggaca 1920
gaggccaagg actgactttg ccaaaaatac aacacagacg atatttaatt cacctcagcc 1980
tgagagggcc tgggggtggga cagggagggc cggcggtctt gacaggggc aggcgcagag 2040
gtcccagccc caggcctggc ctgcctgcc tttctgcctt ttaattttgt atgaggctct 2100
caggtcagct gggagccagt gtgccccaaa gccatgtatt tcagggaacct caggggcacc 2160
tcgggtgcc tagccctccc ccagctccc tgcaccgccc cagaagcagc ccctcgaggc 2220
ctacagagga ggccctcaaag caaccgctg gagccacag cagcctgtg ccttcctccc 2280
cgctcctcc cactgggact ccagcagag cccaccagc agccctggcc cacccccag 2340
cctccagaga agccccgac gggctgtctg ggtgtccgc atccagggtc tggcagagcc 2400
tgtgagatga tgcattgatg cctcccctca ggcaggtctg cagagcccgg cccacctcc 2460
ctgcgccctt gaggggccc agcgtctgca gggtgacgc tgagacagca cactgtctga 2520
ggagtctgag gactgtctc ccacagacc tgcagttagg ggccctccat gcgcagatga 2580
ggggccactg acccacctgc gcttctgctg gaggagggga agctgggccc aaaggcccag 2640
ggaggcagcg tgggtctctg caatgtgggc tgccccctgc acacagggtc cacagggcag 2700
gccttgctgg ggtccagggc tgttgaggga ccccgaggc tgaggagcag ccaggaccgc 2760
cctgtcccca tctcaccda gatcaggaac cagggcctcc ctgttcacgg tgacacaggt 2820
cagggtcag agtgacctc ggctgtcacc tgctcacagg gatgctggtg gctggtgaga 2880
ccccgactg cacacgggaa tgcctaggtc ccttcccag ccagccagct gactgcagg 2940
gcacggggac ctggatagtt aagggtttt ccaaacatgc atccatttac tgacacttcc 3000

tgtccttgtt catggagagc tgttcgctcc tcccagatgg ctteggagge ccgcagggcc 3060
 caccttggac cctggtgacc tcctgtcact cactgaggcc atcagggccc tgccccagge 3120
 ctggacgggc cctccttccc tcctgtgccc cagctgccag gtggccctgg ggaggggtgg 3180
 tgtgtgtgtt ggaaggggtc ctgcaggggg aggaggactt ggagggctct ggggcagctg 3240
~~tcctgaaccg actgacctg aggagggcgc ttagtgetge tttgettttc atcaccgtcc 3300~~
 cgcacagtgg acggagggtc ccggttgctg gtcagggtccc catggettgt tctctggaac 3360
 ctgacttttag atgttttggg atcaggagcc cccaacacag gcaagtccac ccataataa 3420
 ccctgccagt gccaggggtg gctggggact ctggcacagt gatccgggc gccaggacag 3480
 cagcactccc gctgcacaca gacggcctag ggtggcgct cagaccccac cctacgtca 3540
 tctctggaag gggcagccct gactgggtcac tggtcagggc agtggccaag cctgctgtgt 3600
 ccttcctcca caagggtccc ccaccgtca gtgtcagcgg gtgacgtgtg ttcttttgag 3660
 tccttgtatg aataaaaggc tggaaaccta aa 3692

<210> 2
 <211> 495
 <212> DNA
 <213> Homo sapiens

<400> 2
 ctctgtccag gtgagcacct ccgggtctgt cccagaggat acacctgtg ttcagtgag 60
 acagagcaga ggctgatcag ggagactgag gccaccttcc gaggcctggg ggaggacagc 120
 ggctcctttc tgggtcacac actggctgcc aggcacagaa aatttgatga gttttttctg 180
 gagatgtctc cagtagccca gcactctctg acccagctct tctccactc ctacggccgc 240
 ctgtatgcc agcacgccct catattcaat ggctgttct ctcggtgcg agacttctat 300
 ggggaatctg gtgaggggtt ggatgacacc ctggcggatt tctgggcaca gctcctggag 360
 agagtgttcc cgctgtgca ccacagtac agcttcccc ctgactacct gctctgcctc 420
 tcacgcttgg cctcatctac cgatggtct ctgcagccct ttggggactc accccgccgc 480
 ctccgctgc aggtg 495

<210> 3
 <211> 1283
 <212> DNA
 <213> Homo sapiens

<400> 3
 tcggcagatg ccgcctggtc cagctatcgt gctcgggtatt cagttttccg gagcagcgt 60
 ctttctctgg ccgcgggagc ggtccccggg ccgagtaccg gattccccgag tttgggaggc 120
 tctgctttcc tccttaggac ccactttgcc gtctgggggt ggctgcagtt atgtccgcgc 180

tgcgacctct cctgcttctg ctgctgcctc tgtgtcccg tccctggccc ggacccggga 240
 gcgaggcaaaa ggtcaccggg agttgtgcag agacccggca ggtgctgggg gcccggggat 300
 atagcttaaa cctaatecct cccgcctga tctcagggtga gcacctccgg gtctgtcccc 360
 aggagtacac ctgctgttcc agtgagacag agcagaggct gatcagggag actgaggcca 420
~~ccttcggagg cctgggtggag gacagcggct ctttctgggt tcacacactg gctgccaggc~~ 480
 acagaaaatt tgatgataac ccggaccctg ttggctgccc gagcctttgt gcaggggcctg 540
 gagactggaa gaaatgtggt cagcgaagcg cttaagggtc cgggtgtctga aggctgcagc 600
 caggctctga tgcgtctcat cggctgtccc ctgtgccggg gggccccctc acttatgcc 660
 tgcaggggct tctgcctcaa cgtggttcgt ggctgtctca gcagcagggg actggagcct 720
 gactggggca actatctgga tggctcctg atcctggctg ataagctcca gggccccctt 780
 tcccttgagc tgacggccga gtccattggg gtgaagatct cggagggttt gatgtacctg 840
 caggaaaaca gtgcgaaggt gtccgccag gtgtttcagg agtgcgggcc ccccgaccg 900
 gtgcctgcc gcaaccgtcg agccccgcg ccccggaag aggcgggcg gctgtggctg 960
 atggtgaccg aggaggagcg gccacgacg gccgcaggca ccaacctgca ccggtgggtg 1020
 tgggagctcc gcgagcgtct gggccggatg cggggcttct gggcccggt gtccctgacg 1080
 gtgtgcggag actctcgcat ggcagcggac gcctcgctgg aggcggcgcc ctgctggacc 1140
 ggagccgggc ggggcccgtta cttgccgcca gtggctgggg gctccccggc cgagcaggtc 1200
 aacaaccccg agctcaaggt ggacgcctcg ggccccgatg tcccgacacg gcggcgctcg 1260
 ctacagctcc gggcgggcac ggc 1283

<210> 4
 <211> 2312
 <212> DNA
 <213> Homo sapiens

<400> 4
 ccctgccccg cgcgcgaag cggttccgc cctcgcccag cggccaggta gctgcgagga 60
 aacttttgca gcggtgggt agcagcagct ctcttgctcc tcagggccac tgccaggctt 120
 gccgagtcct gggactgctc tcgctccggc tgccactctc ccgcgtctc ctagctccct 180
 gcgaagcagg atggccggga ccgtgcgcac cgcgtgcttg gtggtggcga tgctgctcag 240
 cttggacttc ccgggacagg cgcagcccc gccgcggcg ccggacgcca cctgtcacca 300
 agtcgcctcc ttcttcaga gactgcagcc cggactcaag tgggtgccag aaactccgt 360
 gccaggatca gatattgcaag tatgtctccc taagggccca acatgctgct caagaaagat 420
 ggaagaaaaa taccaactaa cagcagatt gaacatggaa cagctgcttc agtctgcaag 480

tatggagctc aagttcttaa ttattcagaa tgctgcggtt ttccaagagg cctttgaaat	540
tggtgttcgc catgccaaga actacaccaa tgccatgttc aagaacaact acccaagcct	600
gactccacaa gcttttgagt ttgtgggtga atttttcaca gatgtgtctc tctacatctt	660
gggtttctgac atcaatgtag atgacatggt caatgaattg ttgacagcc tgtttccagt	720
catctatacc cagctaata acccaggcct gcctgattca gccttgga tcaatgagt	780
cctccgagga gcaagacgtg acctgaaagt atttgggaat ttccccaagc ttattatgac	840
ccagggtttcc aagtcaactg aagtcactag gatcttctct caggctctga atcttggaat	900
tgaagtgatc aacacaactg atcacctgaa gttcagtaag gactgtggcc gaatgctcac	960
cagaatgtgg tactgtctct actgccaggg actgatgatg gttaaaccct gtggcgggta	1020
ctgcaatgtg gtcattgcaag gctgtatggc aggtgtggtg gagattgaca agtactggag	1080
agaatacatt ctgtcccttg aagaacttgt gaatggcatg tacagaatct atgacatgga	1140
gaacgtactg cttgggtctct tttcaacaat ccatgattct atccagtatg tccagaagaa	1200
tgccaggaaag ctgaccacca ctattggcaa gttatgtgcc cattctcaac aacgccata	1260
tagatctgct tattatctct aagatctctt tattgacaag aaagtattaa aagttgctca	1320
tgtagaacat gaagaaacct tatccagccg aagaaggga ctaattcaga agttgaagtc	1380
tttcatcagc ttctatagtg ctttgctggg ctacatctgc agccatagcc ctgtggcggg	1440
aaacgacacc ctttgctgga atggacaaga actcgtggag agatacagcc aaaaggcagc	1500
aagggaatgga atgaaaaacc agttcaatct ccatgagctg aaaatgaagg gccctgagcc	1560
agtggtcagt caaattattg acaaactgaa gcacattaac cagctcctga gaaccatgtc	1620
tatgccc aaa ggtagagttc tggataaaaa cctggatgag gaagggtttg aaagtggaga	1680
ctgcggtgat gatgaagatg agtgcatggg aggcctctgt gatggaatga taaaagtga	1740
gaatcagctc cgcttctctg cagaactggc ctatgatctg gatgtggatg atgcgcctgg	1800
aaacagtcag caggcaactc cgaaggacaa cgagataagc acctttcaca acctcgggaa	1860
cgttcattcc ccgtgaagc ttctcaccag catggccatc tcgggtgtgt gcttctctct	1920
cctgggtgac tgactgcctg gtgccagca catgtgtgc octacagcac cctgtggtct	1980
tctctgataa agggaaccac tttcttattt tttctattt tttttttttt gttatcctgt	2040
atacctctc cagccatgaa gtagaggact aacctgtgt tatgtttctg aaaatcaaat	2100
ggtatctttt ggaggaagat acatttttagt ggtagcatat agattgtcct ttgcaaaga	2160
aagaaaaaaa accatcaagt tgtgccaaat tattctccta tgtttggctg ctagaacatg	2220
gttaccatgt cttctctct cactccctcc cttctctatg ttctctcttt gcatggattt	2280
ctttgaaaaa aaataaattg ctcaataaa aa	2312

```

<210> 5
<211> 3714
<212> DNA
<213> Homo sapiens

<400> 5
gcctggcacc ggggaccgtt gcctgacgcg aggcccagct ctacttttcg ccccgctct 60
cctccgcttg ctgcctctt ccaccaactc caactccttc tccctccagc tccactcgt 120
agtccccgac tccgccagcc ctgggccgcg tgcgtagcgc ccgcttcccg tccgggtccca 180
aagggtgggaa cgcgtccgcc ccggcccgcg ccatggcacg gttcggcttg cccgcgcttc 240
tctgcaccct ggcagtgtc agcgcgcgcg tgcctggctgc cgagctcaag tcgaaaagtt 300
gctcgggaagt gcgacgtctt tacgtgtcca aaggcttcaa caagaacgat gccccctcc 360
acgagatcaa cgggtgatcat ttgaagatct gtccccaggg ttctacctgc tgcctcctcaag 420
agatggagga gaagtacagc ctgcaaagta aagatgattt caaaagtgtg gtcagcgaac 480
agtgcacatc tttgcaagct gtctttgctt cactgtacaa gaagtttgat gaattcttca 540
aagaactact tgaaaatgca gagaaatccc tgaatgatat gtttgtgaag acatatggcc 600
atattatacat gcaaaattct gagctattta aagatctctt cgtagagttg aaacgttact 660
acgtggtggg aaatgtgaac ctggaagaaa tgctaaatga cttctgggct cgcctcctgg 720
agcggatgtt ccgctggtg aactcccagt accactttac agatgagtat ctggaatgtg 780
tgagcaagta tacggagcag ctgaagccct tcggagatgt ccctcgcaa ttgaagctcc 840
aggttactcg tgcctttgta gcagcccgta ctttcgctca aggccttagcg gttgcgggag 900
atgtcgtgag caaggctctc gtggtaaacc ccacagccca gtgtacccat gccctgttga 960
agatgatcta ctgctccac tgcgggggtc tcgtgactgt gaagccatgt tacaactact 1020
gctcaaacat catgagaggc tgtttggcca accaagggga tctcgatttt gaatggaaca 1080
atctcataga tgcctatgtg atggtggcag agaggctaga gggctccttc aacattgaat 1140
cggtcattga tcccatgat gtgaagattt ctgatgctat tatgaacatg caggataata 1200
gtgttcaagt gtctcagaag gttttccagg gatgtggacc cccaagccc ctcccagctg 1260
gacgaatttc tegtccatc tctgaaagtg ccttcagtgc tgccttcaga ccacatcacc 1320
ccgaggaacg cccaaccaca gcagctggca ctagtttgga ccgactgggt actgatgtca 1380
aggagaaact gaaacaggcc aagaaattct ggtcctccct tccgagcaac gtttgcaacg 1440
atgagaggat ggctgcagga aacggcaatg aggatgactg ttggaatggg aaaggcaaaa 1500
gcaggtacct gtttgacgtg acaggaaatg gattagccaa ccagggaac aaccagagg 1560
tccaggttga caccagcaa ccagacatac tgatccttcg tcaaatcatg gctcttcgag 1620
tgatgaccag caagatgaag aatgcataca atgggaacga cgtggacttc tttgatatca 1680

```

gtgatgaaag tagtggagaa ggaagtggaa gtggctgtga gtatcagcag tgcccttcag 1740
 agtttgacta caatgccact gaccatgctg ggaagagtgc caatgagaaa gccgacagtg 1800
 ctgggtgccc tectggggca caggcctacc tctcactgt cttctgcac tttgttcttg 1860
 ttatgcagag agagtggaga taattctcaa actctgagaa aaagtgttca tcaaaaagtt 1920
~~aaaaggcacc agttatcaat tttctacat cctagtgaat ttgcttttta aatgaatgga 1980~~
 caacaatgta cagtttttac tatgtggcca ctggtttaag aagtgtgac tttgttttct 2040
 cattcagttt tgggaggaaa agggactgtg cattgagtgt gttcctgctc ccccaaacca 2100
 tgttaaactg ggctaacagt gtaggtacag aactatagtt agttgtgcat ttgtgatttt 2160
 atcactctat tatttgtttg tatgtttttt tctcatttct tttgtgggtt tttttttcca 2220
 actgtgatct cgccttggtt cttacaagca aaccagggtc ccttcttggc acgtaacatg 2280
 tacgtatttc tgaaatatta aatagctgta cagaagcagg ttttatttat catgttatct 2340
 tattaaaaga aaaagcccaa aaagcagtaa aatttccatt tctccctgtt attttagttg 2400
 ccttatcttg agagacgtgg aggtgatttt ctttttttta aattattatt aagacagaat 2460
 gtgagggcac aagcaggctt ctgagccact tgtcagattg tattcaaagc atcaatccaa 2520
 gaaggagggt atgtgtactt catttattgg tgatagtgtg aagagactgc agactactgc 2580
 tttgaatgag ttgaattaca taagctaaga tcaatataag tccatttctt gaaccactt 2640
 atacataaaa tgtaaccat ttagaaaaag attctggata tcatccccct tgaaagatag 2700
 aaagcattca ggatgtccca gttatcacat gttcacactt gggtttaggg gtgttttttt 2760
 ttaaaaccag gcaggttagc tagcccaacc tgtgctagt ttcattgtca cactgacct 2820
 atttgaatta atatcctttg ttagagtgtt cgagatttca aaccaatta tgtacaggga 2880
 gctgtctgag agctagccag aactggggta cagcctgggc tcagggaata gctgtcaaca 2940
 ctggggcaaa gtttttgtct gtgcattgtt atctccattt gttttgggat cccagttttt 3000
 gttttaagag agtataaggt gtctcatttg agtcttttct ttacctagcc cctctttatc 3060
 agtaaaacaa aggacttgcc atggttcaca gcaatgtgct acgatccaag atatcagcca 3120
 aggagccac ttaggggaga actaggtgtc cagatttttg tatgtgttgt ttttcttggg 3180
 ggatggggtg ggggtggagt aggtagagct gagaatacta catcttagtg gtgaccttta 3240
 gccacgtggg tgaagtggca aaggccatgg ccatatctgt tgtcccaggc caaagactaa 3300
 caactgcctt ggaatccct tcttgtgtc cttaccaa atgtagtcat aaaactctga 3360
 taatgtaaca aatcacttct aaaggagttc ccagaagtct tcagaaagac taaaattctg 3420
 tctcttcttg ctttagacag ccattaagat cccaactaat tttaccgaac ctaaaacca 3480
 caaagagggt gtttgtgtta ttgttcaatc ttcagttgta agagtaattc tctattttta 3540

tattgaaaca taattacttg atagctcagg gtctacattt cattcaactt tttacaccaa 3600
 attctgcaga gtgggtcaaaa tggaatattg ggggctgttg taaacagagg cttaatttta 3660
 ttagaagtag ccagttattt attaaagcat gatgttaata aaataggcat attc 3714

<210> 6

<211> 3724

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (2877)..(2877)

<223> "n" is A, C, G, or T

<220>

<221> misc_feature

<222> (3380)..(3380)

<223> "n" is A, C, G, or T

<400> 6

gcctggcacc ggggaccgtt gcctgaacgc aggccacagt ctacttttcg ccccgcgctc 60
 cctccgcctg ctgcctctt ccaccaactc caactccttc tccctccagc tccactcgtc 120
 agtccccgac tccgccagcc ctggccccgc tgcgtagcgc ccgcttcccg tccgggtcca 180
 aagggtgggaa cgtgtccgcc ccggccccga ccatggcacg gttcggcttg cccgcgcttc 240
 tctgcaccct ggcagtgtc agcgcgcgcg tcttggtctc cgagctcaag tcgaaaagtt 300
 gctcggaagt gcgacgtctt tacgtgtcca aaggcttcaa caagaacgat gccccctcc 360
 acgagatcaa cgggtgatcat ttgaagatc gtcccaggg ttctacctgc tgctctcaag 420
 agatggagga gaagtacagc ctgcaaagta aagatgattt caaaagtgtg gtcagcgaac 480
 agtgcaatca tttgcaagct gtctttgctt cactgtacaa gaagtttgat gaattcttca 540
 aagaactact tgaaaatgca gagaaatccc tgaatgatat gtttgtgaag acatatggcc 600
 atttatacat gcaaaattct gagctattta aagatctctt cgtagagttg aaacgttact 660
 acgtggtggg aaatgtgaac ctggaagaaa tgctaaatga cttctgggct cgcctcctgg 720
 agcggatggt ccgcctggtg aactcccagt accactttac agatgagtat ctggaatgtg 780
 tgagcaagta tacggagcag ctgaagccct tcggagatgt ccctcgcaaa ttgaagctcc 840
 aggttactcg tgcttttgta gcagcccgta ctttcgctca aggcttagcg gttgcgggag 900
 atgtcgtgag caaggctctc gtggtaaacc ccacagccca gtgtacccat gccctgttga 960
 agatgatcta ctgtccccc tgccgggggc togtgactgt gaagccatgt tacaactact 1020
 gctcaaacat catgagaggc tgtttggcca accaagggga tctcgatttt gaatggaaca 1080

atttcataga	tgctatgctg	atgggtggcag	agaggctaga	gggtcctttc	aacattgaat	1140
cggtcacgga	tcccatcgat	gtgaagat	ctgatgctat	tatgaacatg	caggataata	1200
gtgttcaagt	gtctcagaag	gttttccagg	gatgtggacc	ccccaaagccc	ctcccagctg	1260
gacgaatttc	tcgttccatc	tctgaaagtg	ccttcagtgc	tcgcttcaga	ccacatcacc	1320
ccgaggaacg	ccaaccaca	gcagctggca	ctagtttggg	ccgactgggt	actgatgtca	1380
aggataaact	gaaacaggcc	aagaaattct	ggctcctcct	tccgagcaac	gtttgcaacg	1440
atgagaggat	ggctgcagga	aacggcaatg	aggatgactg	ttggaatggg	aaaggcaaaa	1500
gcagggtacct	gtttgcagtg	acaggaaatg	gattagtcaa	ccagggcaac	aaccagagg	1560
tccaggttga	caccagcaaa	ccagacatac	tgatccttog	tcaaatacatg	gctcttcgag	1620
tgatgaccag	caagatgaag	aatgcataca	atgggaacga	cgtggacttc	tttgatatca	1680
gtgatgaaag	tagtggagaa	ggaagtggaa	gtggctgtga	gtatcagcag	tgcccttcag	1740
agtttgacta	caatgccact	gaccatgctg	ggaagagtgc	caatgagaaa	gccgacagtg	1800
ctggtgtccg	tccctggggca	caggcctacc	tccctactgt	cttctgcatac	ttgttccctgg	1860
ttatgcagag	agagtggaga	taattctcaa	actctgagaa	aaagtgttca	tcaaaaagtt	1920
aaaaggcacc	agttatcact	tttctacat	cctagtgaat	ttgcttttta	aatgaatgga	1980
caacaatgta	cagtttttac	tatgtggcca	ctggtttaag	aagtgtgac	tttgttttct	2040
cattcagttt	tgggaggaaa	agggactgtg	cattgagttg	gttcctgctc	ccccaaacca	2100
tgtaaacgt	ggctaacagt	gtaggtagag	aactatagtt	agttgtgcat	ttgtgatatt	2160
atcactctat	tatttgtttg	tatgtttttt	tctcatttcg	tttggtgggt	tttttttcca	2220
actgtgatct	cgccttggtt	cttacaagca	aaccagggtc	ccttcttggc	acgtaacatg	2280
tacgtatttc	tgaatatatta	aatagctgta	cagaagcagg	ttttatattat	catgttatct	2340
tattaaaaga	aaaagcccaa	aaagcagtaa	aatttccatt	tctccctgtt	attttagttg	2400
ccttatcttg	agagacgtgg	aggatgattt	ctttttttaa	attattatta	agacagaatg	2460
tgaaggcaca	agcaggcttc	tgagccactt	gtcagattgt	attcaaagca	tcaatccaag	2520
aggaggttat	gtgtacttca	tttattgggt	atagttggaa	gagactgcag	actactgctt	2580
tgaatgagtt	gaattacata	agctaagatc	actataaggt	ccatttcttg	aaccacttta	2640
tacataaaat	gtaaccatt	tagaaaaaga	ttctggatat	catccccct	tgaagatag	2700
aaagcattca	ggatgtccca	gttatcacat	gttcacactt	gggttttaggg	gtgttttttt	2760
ttaaaaccag	gcaggtttag	tagccacccc	tgtgctagtt	ttcatgttca	cgtgaccct	2820
atttgaatta	atatcctttg	ttagagtggg	cgagatttca	aaccaatta	tgtacangga	2880
gctgtctgag	agctagccag	aactggggta	cagcctgggc	tcagggaata	gctgtcaaca	2940


```

ctcgggcaaa gtttttgtct gcagcaacgt gtatcaccat ttgttttggg atccagtttt 3000
tgttttaaga gagtataagg tgtctcattt gagtcttttt cttacctagc cccctcttat 3060
cagtaaaaca aaggacttgc catggttcac agcaatgtgc tacgatccaa gatatcaacc 3120
aaggagccca cttaggggag aactagggtgt ccagattttt gtatgtgttg tttttcttgg 3180
gggatggggt ggggtgggag taggtagagc tgagaatact acatcttagt ggtgaccttt 3240

agccacgtgg gtgaagtggc aaaggccatg gccatatctg ttgtcccagg ccaaagacta 3300
acaactgcct tgggaatccc ttccttgtgt ccttaccaaa tgatagctca taaaactctg 3360
ataatgtaac aaatcacttn caaaggagtt ccagaagtc ttcagaaaga ctaaaattct 3420
gtctcttccct gcttttagaca gccatlaaga tcccaactaa ttttaccgaa cctaaaaccc 3480
acaaagaggt tgtttgtgtt attgttcaat cttcagttgt aagagtaatt ctctattttt 3540
atattgaaac ataattactt gatagctcag ggtctacact tcattcaact ttttacacca 3600
aattctgcag agtgggtcaaa atggaatatt gggggctgtt gtaaacagag gcttaatttt 3660
attagaagta gccagttatt tattaagca tgatgttaat aaaataggca tattccaaaa 3720
aaaa 3724

```

```

<210> 7
<211> 1722
<212> DNA
<213> Homo sapiens

```

```

<400> 7
accatggcac ggttcggctt gcccgcgctt ctctgcaccc tggcagtgct cagcgccgcg 60
ctgctggctg ccgagctcaa gtcgaaaagt tgctcggaag tgcgacgtct ttacgtgtcc 120
aaaggttca acaagaacga tgccccctc cagcagatca acggtgatca tttgaagatc 180
tggtcccagg gttctacctg ctgctctcaa gagatggagg agaagtacag cctgcaaagt 240
aaagatgatt tcaaaagtgt ggtcagcgaa cagtgcacac atttgcaagc tgtcttttgt 300
tcacgttaca agaagtttga tgaattcttc aaagaactac ttgaaaatgc agagaaatcc 360
ctgaatgata tgtttgtgaa gacatatggc catttataca tgcaaaatcc tgagctattt 420
aaagatctct tcgtagagtt gaaacgttac tacgtggtgg gaaatgtgaa cctggaagaa 480
atgctaaatg acttctgggc tcgcctcctg gagcggatgt tccgcctggt gaactccag 540
taccacttta cagatgagta tctggaatgt gtgagcaagt atacggagca gctgaagccc 600
ttcggagatg tccctcgcaa attgaagctc caggttactc gtgcttttgt agcagcccg 660
actttcgctc aaggcttagc ggttgcgga gatgtcgtga gcaaggctc cgtggtaaac 720
cccacagccc agtgtacca tgccctgttg aagatgatct actgtccca ctgccggggt 780
ctcgtgactg tgaagccatg ttacaactac tgctcaaaac tcatgagagg ctgtttggcc 840

```

```

aaccaagggg atctcgattt tgaatggaac aatttcatag atgctatgct gatggtggca 900
gagaggctag agggtccttt caacattgaa toggatcatgg atcccatcga tgtgaagatt 960
tctgatgcta ttatgaacat gcaggataat agtgttcaag tgtctcagaa ggttttccag 1020
ggatgtggac cccccaagcc cctcccagct ggacgaatct ctcgttccat ctctgaaagt 1080
gctttcagtg ctcgcttcag accacateac cccgaggaac gcccacccac agcagctggc 1140
actagtttgg accgactggg tactgatgtc aaggagaaac tgaaacaggc caagaaattc 1200
tggctcctcc ttccgagcaa cgtttgcaac gatgagagga tggctgcagg aaacggcaat 1260
gaggatgact gttggaatgg gaaaggcaaa agcaggatcc tgtttgcagt gacaggaaat 1320
ggattagcca accagggcaa caaccagag gtccaggttg acaccagcaa accagacata 1380
ctgatccttc gtcaaatcat ggctcttcga gtgatgacca gcaagatgaa gaatgcatac 1440
aatgggaacg acgtggactt ctttgatata agtgatgaaa gtagtggaga aggaagtggg 1500
agtggctgtg agtatcagca gtgcccttca gagtttgact acaatgccac tgaccatgct 1560
gggaagagtg ccaatgagaa agccgacagt gctggtgtcc gtccctggggc acaggcctac 1620
ctcctcactg tcttctgcat cttgttcttg gttatgcaga gagagtggag ataattctca 1680
aactctgaga aaaagtgttc atcaaaaagt taaaaggcac ca 1722

```

```

<210> 8
<211> 2625
<212> DNA
<213> Homo sapiens

```

```

<400> 8
gtcttccacg tctgcagctc agccagggcg cgcagggcga gtgggggtcca ctggcgggta 60
aaggggacca ggacggcgag gatggacgca cagacctggc ccgtgggctt tcgctgcctc 120
ctccttcttg ccttggttgg gtccgccgcg agcgagggcg tgcagacctg cgaagaagtt 180
cggaaacttt tccagtggcg gctgctggga gctgtcaggg ggctgccgga ttcgccgcgg 240
gcaggacctg atcttcaggt ttgcatatcc aaaaagccta catgttgcac caggaagatg 300
gaggagagat atcagattgc ggctcgcag gatatgcagc agtttcttca aacgtccagc 360
tctacattaa agtttctaata atctcgaaat gcggctgctt ttcaagaaac ccttgaaact 420
ctcatcaaac aagcagaaaa ttacaccagt atactttttt gcagtacctt caggaacatg 480
gcttggaggg ctgctgcttc ggttcaggag ttcttctactg atgtggggct gtatttattt 540
ggtgcggatg ttaatcctga agaatttgta aacagatttt ttgacagtct ttttctctg 600
gtctacaacc acctcattaa ccctggtgtg actgacagtt ccctggaata ctcagaatgc 660
atccggatgg ctgccgggga tgtgagtgca ttgggtaata tccccaaag agtaatggga 720

```

cagatgggga ggtccctgct gccacgccc acttttctgc aggcactcaa tctgggcatt	780
gaagtcacaa acaccacaga ctatctgcac ttctccaaag agtgcagcag agccctctctg	840
aagatgcaat actgcccga ctgccaaagg ctggcgctca ctaagccttg tatgggatac	900
tgccctcaatg tcatgcgagg ctgcctggcg cacatggcgg agcttaatcc acactggcat	960
gcataatccc ggtcgttgga agaactctcg gatgcaatgc atggaacata cgacattgga	1020
cacgtgctgc tgaactttca cttgcttggt aatgatgctg tggtacaggc tcacctcaat	1080
ggacaaaaat tattggaaca ggtaaatagg atttgtggcc gccctgtaag aacaccaca	1140
caaagcccc gttgttcttt tgatcagagc aaagagaagc atggaatgaa gaccaccaca	1200
aggaacagtg aagagacgct tgccaacaga agaaaagaat ttatcaacag ccttcgactg	1260
tacaggteat tctatggagg tctagctgat cagctttgtg ctaatgaatt agctgtgca	1320
gatggacttc cctgctggaa tggagaagat atagtaaaaa gttatactca gcgtgtggtt	1380
ggaaatggaa tcaaagcccc gtctggaaat cctgaagtca aagtcaaagg aattgatcct	1440
gtgataaatc agattattga taaactgaag catgttgttc agttgttaca gggtagatca	1500
cccaaacctg acaagtggga acttcttcag ctgggcagtg gtggaggcat ggttgaacaa	1560
gtcagtgggg actgtgatga tgaagatggt tgcgggggat caggaagtgg agaagtcaag	1620
aggacactga agatcacaga ctggatgcca gatgatatga acttcagtga tgtaaagcaa	1680
atccatcaaa cagacactgg cagtacttta gacacaacag gagcaggatg tgcagtggcg	1740
actgaatcta tgacattcac tctgataagt gtggtgatgt tacttcccg gatttggtaa	1800
ctgaactctt ctgtcctgac ataccttact gaagtctcga tttcttctct ctctgcatat	1860
gcctggaata agagatcctt tttcaatgta acaattatat ttatgaaaag atatgttaca	1920
ctaacttctc agaagccaag ctgaaatatt cataaagtc ctaaaactca acgtttaaat	1980
gacacacttt aaaaatatgt cttttttcaa tctaactgaa aaccttctta acttctaata	2040
tattaaatct gaagatgtga agggcacaga agtgactttg aataagaaga atttagtgta	2100
tctgtaattt tattatcaat tccaagcccc ttcttttcta aattaaaaat gttttcattt	2160
gaaagtgtat ttgccagaca atgaaaacag tatgcagtat ttcttaaagt attgaaatta	2220
gaatatcatg aaataaatca aaacatacaa tggcaagtag tatgcatgca tattcaagag	2280
actcttccat ttttgcaagc tgtagaagga aatgtctgaa tgtctataag ttatggggta	2340
gattcttgag aagcatttca tataatttca ctgaagaacc ttgataattt tgaccactg	2400
taacttagcc actgatgaac cttaaagctg agtattttat taacacctga tttgtattct	2460
attatattca aatgcatct ttggtattgt gcctctgctc ccactctctt ctttgccctca	2520
tagatttagc tatgttggga agcacatgct tgctctagga atatctccaa taaagctgtt	2580

aactatattgg tggaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaa 2625

<210> 9

<211> 1938

<212> DNA

<213> Homo sapiens

<400> 9

gtcttccacg tctgcagctc agccagggcg cgcagggcga gtgggggtcca ctggcgggta 60
aaggggacca ggacggcgag gatggacgca cagacctggc ccgtgggctt tcgctgcctc 120
ctccttcttg ccctgggttg gtccgcccg cgcagggcg tgacagacct cgaagaagtt 180
cggaaacttt tccagtggcg gctgctggga gctgtcaggg ggctgccgga ttcgcccggg 240
gcaggacctg atcttcaggt ttgcataacc aaaaagccta catgttgac caggaagatg 300
gaggagagat atcagattgc ggctcgccag gatatgcagc agtttcttca aacgtccagc 360
tctacattaa agtttcta atctcgaaat gcggctgctt ttcaagaaac ccttgaaact 420
ctcatcaaac aagcagaaaa ttacaccagt ataacttttt gcagtacct caggaacatg 480
gccttggagg ctgctgcttc ggctcaggag ttcttcaact atgtggggct gtatttat 540
gggtcgggatg ttaatcctga agaatttgta aacagatttt ttgacagtct ttttccctctg 600
gtctacaacc acctcattaa ccctgggttg actgacagtt ccctggaata ctcagaatgc 660
atccggatgg ctgcgggga tgtgagtcca ttttgtaata ttcccaaag agtaatggga 720
cagatgggga ggtccctgct gccagccgc acttttctgc aggcactcaa tctgggcatt 780
gaagtcacat acaccacaga ctatctgcac ttcttcaaag agtgacagc agccctcctg 840
aagatgcaat actgcccgca ctgccaggc ctggcgctca ctaagccttg tatgggatac 900
tgccctcaatg tcatgcgagg ctgcctggcg cacatggcgg agcttaatcc aactggcat 960
gcatatatcc ggtcggttga agaactctcg gatgcaatgc atggaacata cgacattgga 1020
cacgtgctgc tgaactttca cttgctgtt aatgatgctg tgttacaggc tcacctcaat 1080
ggacaaaaat tattggaaca ggtaaatagg atttgtggcc gccctgtaag aacaccaca 1140
caaagcccc gtgttcttt tgatcagagc aaagagaagc atggaatgaa gaccaccaca 1200
aggaacagtg aagagacgct tgccaacaga agaaaagaat ttatcaacag ccttcgactg 1260
tacaggatcat tctatggagg tctagctgat cagctttgtg ctaatgaatt agctgctgca 1320
gatggacttc cctgctggaa tggagaagat atagtaaaaa gttatactca gcgtgtggtt 1380
ggaaatggaa tcaaagccca gtctggaat cctgaagtca aagtcaaagg aattgatcct 1440
gtgataaatc agattattga taaactgaag catgtgttgc agttgttaca gggtagatca 1500
cccaaacctg acaagtggga acttcttcag ctgggcagtg gtggaggcat ggttgaacaa 1560
gtcagtgggg actgtgatga tgaagatggg tgcgggggat caggaagtgg agaagtcaag 1620

```

aggacactga agatcacaga ctggatgcc a gatgatatga acttcagtga tgtaaagcaa 1680
atccatcaaa cagacactgg cagtacttta gacacaacag gagcaggatg tgcagtggcg 1740
actgaatcta tgacattcac tctgataagt gtggtgatgt tacttcccg gatttggtaa 1800
ctgaactctt ctgtcctgac ataccttact gaagtctcga tttcttctct ctctgcata 1860
gcttgggaata agagatcctt tttcaatgta acaattatat ttatgaaaag atatgttaca 1920
ctaacttcca gaagccaa 1938

```

```

<210> 10
<211> 2644
<212> DNA
<213> Homo sapiens

```

```

<400> 10
agcggcgct gaattctagc caggatttct ctcttctat ttcaggagga ctctcacagg 60
ctcccacagc ctgtgttaag ctgaggtttc ccttagatct cgtatatccc caacacatac 120
ctccacgcac acacatcccc aagaacctcg agctcacacc aacagacaca cgcgcgcata 180
cacactcgct ctgccttgtc catctccctc cggggggagc cggcgcgcgc toccaccttt 240
gccgcacact cggcgagacc gagcccgag cgctccagga ttctgaggct cggaactcgg 300
attgcagctc tgaaccccca tgggtggttt ttaaactctt cttttccttc tcttctcgt 360
tttgattgca ccgtttccat ctgggggcta gaggagcaag gcagcagcct toccagccag 420
cccttggttg cttgccatcg tccatctggc ttataaaagt ttgctgagcg cagtccagag 480
ggctgcgctg ctgcctccct cggctggcag aaggggggtga cgtgggcag cggcgaggag 540
cgcgcgcctg cctctggcgg gctttcggt tgaggggcaa ggtgaagagc gcaccggcgg 600
tggggtttac cgagctggat ttgtatgttg caccatgcct tcttggatcg gggctgtgat 660
tcttccctc ttggggctgc tgctctccct ccccgccggg gcggatgtga aggctcggag 720
ctgcggagag gtccgccagg cgtacggtgc caagggatc agcctggcgg acatccccta 780
ccaggagatc gcaggggaac acttaagaat ctgtcctcag gaatatacat gctgcaccac 840
agaaatggaa gacaagttaa gccaacaaag caaactcgaa tttgaaaacc ttgtggaaga 900
gacaagccat tttgtgcgca ccacttttgt gtccaggcat aagaaatttg acgaattttt 960
ccgagagctc ctggagaatg cagaaaagtc actaaatgat atgtttgtac ggacctatgg 1020
catgctgtac atgcagaatt cagaagtctt ccaggacctc ttcacagagc tgaaaaggta 1080
ctacactggg ggtaatgtga atctggagga aatgctcaat gacttttggg ctcggtcct 1140
ggaacggatg tttcagctga taaacctca gtatcacttc agtgaagact acctggaatg 1200
tgtgagcaaa tacactgacc agtcaagcc atttggagac gtgccccgga aactgaagat 1260

```

tcagggttacc cgcgccttca ttgctgccag gacctttgtc caggggctga ctgtgggcag 1320
 agaagttgca aaccgagttt ccaaggctag cccaaccca ggggtgtatcc gtgccctcat 1380
 gaagatgctg tactgcccac actgtcgggg gcttcccact gtgaggccct gcaacaacta 1440
 ctgtctcaac gtcctgaagg gctgcttggc aaatcaggct gacctcgaca cagagtggaa 1500
 tctgtttata gatgcaatgc tcttggtggc agagcgactg gaggggcat tcaacattga 1560
 gtcggctcatg gaccgatag atgtcaagat ttctgaagcc attatgaaca tgcaagaaaa 1620
 cagcatgcag gtgtctgcaa aggtctttca gggatgtggt cagcccaaac ctgctccagc 1680
 cctcagatct gcccgctcag ctctgaaaa ttttaataca cgtttcaggc cctacaatcc 1740
 tgaggaaaga ccaacaactg ctgcaggcac aagcttggac cggctgggtca cagacataaa 1800
 agagaaattg aagctctcta aaaaggctct gtcagcatta ccttacctta tctgcaagga 1860
 cgagagcgtg acagcgggca cgtccaacga ggaggaatgc tggaaacggc acagcaaagc 1920
 cagatacttg cctgagatca tgaatgatgg gctcaccaac cagatcaata atcccgaggt 1980
 ggatgtggac atcactcggc ctgacacttt catcagacag cagattatgg ctctccgtgt 2040
 gatgaccaac aaactaaaaa acgctacaa tggcaatgat gtcaatttcc aggacacaag 2100
 tgatgaatcc agtgggtcag ggagtggcag tgggtgcatg gatgacgtgt gtcccacgga 2160
 gtttgagttt gtcaccacag agggcccccgc agtggatccc gaccggagag aggtggactc 2220
 ttctgcagcc cagcgtggcc actccctgct ctctggtct ctacctgca ttgtcctggc 2280
 actgcagaga ctgtgcagat aatcttgggt ttttggtcag atgaaactgc attttagcta 2340
 tctgaatggc caactcactt cttttcttac actcttggac aatggaccat gccacaaaaa 2400
 cttaccgttt tctatgagaa gagagcagta atgcaatctg cctccctttt tgttttccca 2460
 aagagtaccg ggtgccagac tgaactgctt cctctttcct tcagctatct gtggggacct 2520
 tgtttattct agagagaatt ctactcaaa tttttcgtac caggagattt tcttaccttc 2580
 atttgctttt atgctgcaga agtaaaggaa tctcacgttg tgagggtttt tttttttctc 2640
 attt 2644

<210> 11
 <211> 2760
 <212> DNA
 <213> Homo sapiens

<400> 11
 ccaggatttc tctcttcta tttcaggagg actctcacag gctccacag cctgtgttaa 60
 gctgaggttt cccctagatc tcgtatatcc ccaacacata cctccacgca cacacatccc 120
 caagaacctc gagctcacac caacagacac acgcgcgcac acacactcgc tctcgttgt 180
 ccatctccct ccggggggag ccggcgcgcg cttccacctt tgccgcacac tccggcgagc 240

cgagcccgc ggcctccagg attctgccc tcggaactcg gattgcagct ctgaaccccc 300
 atgggtgggtt tttaaacact tcttttcctt ctcttcctcg ttttgattgc accgtttcca 360
 tctgggggct agaggagcaa ggcagcagcc ttcccagcca gcccttggtg gcttgccatc 420
 gtccatcttg cttataaaaag tttgctgagc gcagtcagga gggctgcgct gctcgtcccc 480
~~tgggtggca gaagggggtg aegctgggca gggggagga ggcgcgcgct gcctctggcg~~ 540
 ggctttcggc ttgaggggca aggtgaagag cgcaccggcc gtggggttta ccgagctgga 600
 tttgtatgtt gcaccatgcc ttcttgatc ggggctgtga ttcttccct cttggggctg 660
 ctgctctccc tccccgccg ggcggatgtg aaggctcgga gctgcggaga ggtccgccag 720
 gcgtacgggtg ccaagggatt cagcctggcg gacatccct accaggagat cgcaggggaa 780
 cacttaagaa tctgtcctca ggaatataca tgctgcacca cagaaatgga agacaagtta 840
 agccaacaaa gcaaactcga atttgaaaac cttgtggaag agacaagcca ttttgtgcgc 900
 accacttttg tgtccaggca taagaaattt gacgaatttt tccgagagct cctggagaat 960
 gcagaaaagt cactaaatga tatgtttgta cggacctatg gcctgctgta catgcagaat 1020
 tcagaagtct tccaggacct cttcacagag ctgaaaaggt actacactgg gggtaatgtg 1080
 aatctggagg aaatgctcaa tgacttttgg gctcggctcc tggaacggat gtttcagctg 1140
 ataaacctc agtatcactt cagtgaagac tacctggaat gtgtgagcaa atacactgac 1200
 cagctcaagc catthggaga cgtgccccg aaactgaaga ttcaggttac ccgcgccttc 1260
 attgctgcca ggacctttgt ccaggggctg actgtgggca gagaagttgc aaaccgagtt 1320
 tccaaggcca gcccaacccc aggggtgatc cgtgccctca tgaagatgct gtactgcccc 1380
 tactgtcggg ggcttccac tgtgaggccc tgcaacaact actgtctcaa cgtcatgaag 1440
 ggctgcttg caaatcaggc tgacctcgac acagagtgga atctgtttat agatgcaatg 1500
 ctcttggtg cagagcgact ggaggggcca ttcaacattg agtcggtcat ggacccgata 1560
 gatgtcaaga tttctgaagc cattatgaac atgcaagaaa acagcatgca ggtgtctgca 1620
 aaggcttttc agggatgtgg tcagcccaaa cctgtccag cctcagatc tgcccgctca 1680
 gctcctgaaa attttaatac acgtttcagg ccctacaatc ctgaggaaag accaacaact 1740
 gctgcaggca caagcttgga ccggctggtc acagacataa aagagaaatt gaagctctct 1800
 aaaaaggctc ggtcagcatt accctacact atctgcaagg acgagagcgt gacagcgggc 1860
 acgtccaacg agggaggaatg ctggaacggg cacagcaaag ccagatactt gcctgagatc 1920
 atgaatgatg ggctcaccaa ccagatcaat aatcccaggg tggatgtgga catcactcgg 1980
 cctgacactt tcatcagaca gcagattatg gctctccgtg tgatgaccaa caaactaaaa 2040
 aacgcctaca atggcaatga tgtcaatttc caggacacaa gtgatgaatc cagtggctca 2100

gggagtgggca gtgggtgcat ggatgacgtg tgtcccaagg agtttgagtt tgtcaccaca 2160
 gagggccccc cagtggatcc cgaccggaga gaggtggact cttctgcagc ccagcgtggc 2220
 cactccctgc tctcctggtc tctcacctgc attgtcctgg cactgcagag actgtgcaga 2280
 taatcttggg tttttggcca gatgaaactg cattttagct atctgaatgg ccaactcact 2340
~~tcttttctta cactcttggg caatggacca tgcacaaaa acttaaccgtt ttctatgaga 2400~~
 agagagcagt aatgcaatct gctccctttt ttgttttccc aaagagtacc gggtgccaga 2460
 ctgaactgct tcctctttcc ttcagctatc tgtggggacc ttgtttattc tagagagaat 2520
 tcttactcaa atttttcgta ccaggagatt ttcttacctt catttgcttt tatgctgcag 2580
 aagtaaagga atctcacgtt gtgaggggtt ttttttttct catttaaaat aaaaaaggaa 2640
 gaaagaaaat aattttcctt gtaaaatcgg gccaaacccc aagacagcta catthttcaac 2700
 aaaaaagcaa acagagaaaa ataatgaac tttaacactg taagttcagc attgacagcc 2760

<210> 12
 <211> 1799
 <212> DNA
 <213> Homo sapiens

<400> 12
 gccgtggggt ttaccgagct ggatttgtat gttgcaccat gccttcttgg atcggggctg 60
 tgattcttcc cctcttgggg ctgctgctct ccctccccgc cggggcggat gtgaaggctc 120
 ggagctgcgg agaggtccgc caggcgtacg gtgccaaagg attcagcctg gccgacatcc 180
 cctaccagga gatgcaggg gaacaactaa gaatctgtcc tcaggaatat acatgctgca 240
 ccacagaaat ggaagacaag ttaagccaac aaagcaaact cgaatttgaa aaccttgtgg 300
 aagagacaag ccattttgtg cgcaccactt ttgtgtccag gcataagaaa ttgacgaat 360
 ttttccgaga gctcctggag aatgcagaaa agtcactaaa tgatatgttt gtacggacct 420
 atggcatgct gtacatgcag aattcagaag tcttcagga cctcttcaca gagctgaaaa 480
 ggtactacac tgggggtaat gtgaatctgg aggaaatgct caatgacttt tgggctcggc 540
 tcctggaacg gatgtttcag ctgataaacc ctacgtatca cttcagtga gactacctgg 600
 aatgtgtgag caaatacact gaccagctca agccatttgg agacgtgccc cggaaactga 660
 agattcaggc taccgcgcc ttcattgtg ccaggacctt tgtccagggg ctgactgtgg 720
 gcagagaagt tgcaaacga gtttccaagg tcagcccaac ccagggtgt atcgtgccc 780
 tcatgaagat gctgtactgc ccatactgtc gggggcttcc cactgtgagg ccctgcaaca 840
 actactgtct caacgtcatg aagggtgct tggcaaatca ggctgacctc gacacagagt 900
 ggaatctgtt tatagatgca atgctcttgg tggcagagcg actggagggg ccattcaaca 960

ttgagtcggt catggacccg atagatgtca agatttctga agccattatg aacatgcaag 1020
 aaaaacagcat gcaggtgtct gcaaaggctt ttcagggatg tggtcagccc aaacctgtctc 1080
 cagccctcag atctgcccgc tcagctcctg aaaattttta tacacgtttc aggccttaca 1140
 atcttgagga aagaccaaca actgctgcag gcacaagctt ggaccggctg gtcacagaca 1200
 taaaagagaa attgaagctc tctaaaaagg tctggtcagc attaccctac actatctgca 1260

aggacgagag cgtgacagcg ggcacgtcca acgaggagga atgctggaac gggcacagca 1320
 aagccagata cttgcctgag atcatgaatg atgggctcac caaccagatc aacaatcccg 1380
 aggtgggatgt ggacatcaact cggcctgaca ctttcatcag acagcagatt atggtctctc 1440
 gtgtgatgac caacaaacta aaaaacgcct acaatggcaa tgatgtcaat ttccaggaca 1500
 caagtgatga atccagtggc tcaggagtg gcatgggtg catggatgac gtgtgtccca 1560
 cggagtttga gtttgtcacc acagaggccc ccgcagtga tcccgaccgg agagaggtgg 1620
 actcttctgc agcccagcgt ggccactccc tgcctcctg gtctctcacc tgcattgtcc 1680
 tggcactgca gagactgtgc agataatctt gggtttttgg tcagatgaaa ctgcatttta 1740
 gctatctgaa tggccaactc acttcttttc ttacactctt ggacaatgga ccatgccac 1799

<210> 13
 <211> 558
 <212> PRT
 <213> Homo sapiens

<400> 13

Met Glu Leu Arg Ala Arg Gly Trp Trp Leu Leu Cys Ala Ala Ala Ala
 1 5 10 15

Leu Val Ala Cys Ala Arg Gly Asp Pro Ala Ser Lys Ser Arg Ser Cys
 20 25 30

Gly Glu Val Arg Gln Ile Tyr Gly Ala Lys Gly Phe Ser Leu Ser Asp
 35 40 45

Val Pro Gln Ala Glu Ile Ser Gly Glu His Leu Arg Ile Cys Pro Gln
 50 55 60

Gly Tyr Thr Cys Cys Thr Ser Glu Met Glu Glu Asn Leu Ala Asn Arg
 65 70 75 80

Ser His Ala Glu Leu Glu Thr Ala Leu Arg Asp Ser Ser Arg Val Leu
 85 90 95

Gln Ala Met Leu Ala Thr Gln Leu Arg Ser Phe Asp Asp His Phe Gln
 100 105 110

His Leu Leu Asn Asp Ser Glu Arg Thr Leu Gln Ala Thr Phe Pro Gly
 115 120 125

Ala Phe Gly Glu Leu Tyr Thr Gln Asn Ala Arg Ala Phe Arg Asp Leu
 130 135 140

Tyr Ser Glu Leu Arg Leu Tyr Tyr Arg Gly Ala Asn Leu His Leu Glu
 145 150 155 160

Glu Thr Leu Ala Glu Phe Trp Ala Arg Leu Leu Glu Arg Leu Phe Lys
 165 170 175

Gln Leu His Pro Gln Leu Leu Leu Pro Asp Asp Tyr Leu Asp Cys Leu
 180 185 190

Gly Lys Gln Ala Glu Ala Leu Arg Pro Phe Gly Glu Ala Pro Arg Glu
 195 200 205

Leu Arg Leu Arg Ala Thr Arg Ala Phe Val Ala Ala Arg Ser Phe Val
 210 215 220

Gln Gly Leu Gly Val Ala Ser Asp Val Val Arg Lys Val Ala Gln Val
 225 230 235 240

Pro Leu Gly Pro Glu Cys Ser Arg Ala Val Met Lys Leu Val Tyr Cys
 245 250 255

Ala His Cys Leu Gly Val Pro Gly Ala Arg Pro Cys Pro Asp Tyr Cys
 260 265 270

Arg Asn Val Leu Lys Gly Cys Leu Ala Asn Gln Ala Asp Leu Asp Ala
 275 280 285

Glu Trp Arg Asn Leu Leu Asp Ser Met Val Leu Ile Thr Asp Lys Phe
 290 295 300

Trp Gly Thr Ser Gly Val Glu Ser Val Ile Gly Ser Val His Thr Trp
 305 310 315 320

Leu Ala Glu Ala Ile Asn Ala Leu Gln Asp Asn Arg Asp Thr Leu Thr
 325 330 335

Ala Lys Val Ile Gln Gly Cys Gly Asn Pro Lys Val Asn Pro Gln Gly
 340 345 350

Pro Gly Pro Glu Glu Lys Arg Arg Arg Gly Lys Leu Ala Pro Arg Glu
 355 360 365

Arg Pro Pro Ser Gly Thr Leu Glu Lys Leu Val Ser Glu Ala Lys Ala
 370 375 380

Gln Leu Arg Asp Val Gln Asp Phe Trp Ile Ser Leu Pro Gly Thr Leu
 385 390 395 400

Cys Ser Glu Lys Met Ala Leu Ser Thr Ala Ser Asp Asp Arg Cys Trp
 405 410 415

Asn Gly Met Ala Arg Gly Arg Tyr Leu Pro Glu Val Met Gly Asp Gly
 420 425 430

Leu Ala Asn Gln Ile Asn Asn Pro Glu Val Glu Val Asp Ile Thr Lys
 435 440 445

Pro Asp Met Thr Ile Arg Gln Gln Ile Met Gln Leu Lys Ile Met Thr
 450 455 460

Asn Arg Leu Arg Ser Ala Tyr Asn Gly Asn Asp Val Asp Phe Gln Asp
 465 470 475 480

Ala Ser Asp Asp Gly Ser Gly Ser Gly Ser Gly Asp Gly Cys Leu Asp
 485 490 495

Asp Leu Cys Gly Arg Lys Val Ser Arg Lys Ser Ser Ser Ser Arg Thr
 500 505 510

Pro Leu Thr His Ala Leu Pro Gly Leu Ser Glu Gln Glu Gly Gln Lys
 515 520 525

Thr Ser Ala Ala Ser Cys Pro Gln Pro Pro Thr Phe Leu Leu Pro Leu
 530 535 540

Leu Leu Phe Leu Ala Leu Thr Val Ala Arg Pro Arg Trp Arg
 545 550 555

<210> 14
 <211> 579
 <212> PRT
 <213> Rattus norvegicus

<400> 14

Met Ser Ala Val Arg Pro Leu Leu Leu Leu Leu Leu Pro Leu Cys Pro
 1 5 10 15

Gly Pro Gly Pro Gly His Gly Ser Glu Ala Lys Val Val Arg Ser Cys
 20 25 30

Ala Glu Thr Arg Gln Val Leu Gly Ala Arg Gly Tyr Ser Leu Asn Leu
 35 40 45

~~Ile Pro Pro Ser Leu Ile Ser Gly Glu His Leu Gln Ile Cys Pro Gln~~
 50 55 60

Glu Tyr Thr Cys Cys Ser Ser Glu Thr Glu Gln Lys Leu Ile Arg Asp
 65 70 75 80

Ala Glu Val Thr Phe Arg Gly Leu Val Glu Asp Ser Gly Ser Phe Leu
 85 90 95

Ile His Thr Leu Ala Ala Arg His Arg Lys Phe Asn Glu Phe Phe Arg
 100 105 110

Glu Met Leu Ser Ile Ser Gln His Ser Leu Ala Gln Leu Phe Ser His
 115 120 125

Ser Tyr Gly Arg Leu Tyr Ser Gln His Ala Val Ile Phe Asn Ser Leu
 130 135 140

Phe Ser Gly Leu Arg Asp Tyr Tyr Glu Lys Ser Gly Glu Gly Leu Asp
 145 150 155 160

Asp Thr Leu Ala Asp Phe Trp Ala Gln Leu Leu Glu Arg Ala Phe Pro
 165 170 175

Leu Leu His Pro Gln Tyr Ser Phe Pro Pro Asp Phe Leu Leu Cys Leu
 180 185 190

Thr Arg Leu Thr Ser Thr Ala Asp Gly Ser Leu Gln Pro Phe Gly Asp
 195 200 205

Ser Pro Arg Arg Leu Arg Leu Gln Ile Thr Arg Ala Leu Val Ala Ala
 210 215 220

Arg Ala Leu Val Gln Gly Leu Glu Thr Gly Arg Asn Val Val Ser Glu
 225 230 235 240

Ala Leu Lys Val Pro Met Leu Glu Gly Cys Arg Gln Ala Leu Met Arg
 245 250 255

Leu Ile Gly Cys Pro Leu Cys Arg Gly Val Pro Ser Leu Met Pro Cys

260	265	270
Arg Gly Phe Cys Leu Asn Val	Ala His Gly Cys Leu Ser Ser Arg Gly	
275	280	285
Leu Glu Pro Glu Trp Gly Gly Tyr Leu Asp Gly Leu Leu Leu Leu Ala		
290	295	300
Glu Lys Leu Gln Gly Pro Phe Ser Phe Glu Leu Ala Ala Glu Ser Ile		
305	310	315 320
Gly Val Lys Ile Ser Glu Gly Leu Met His Leu Gln Glu Asn Ser Val		
	325	330 335
Lys Val Ser Ala Lys Val Phe Gln Glu Cys Gly Thr Pro His Pro Val		
	340	345 350
Gln Ser Arg Asn Arg Arg Ala Pro Ala Pro Arg Glu Glu Thr Ser Arg		
	355	360 365
Ser Trp Arg Ser Ser Ala Glu Glu Glu Arg Pro Thr Thr Ala Ala Gly		
	370	375 380
Thr Asn Leu His Arg Leu Val Trp Glu Leu Arg Glu Arg Leu Ser Arg		
	385	390 395 400
Val Arg Gly Phe Trp Ala Gly Leu Pro Val Thr Val Cys Gly Asp Ser		
	405	410 415
Arg Met Ala Ala Asp Leu Ser Gln Glu Ala Ala Pro Cys Trp Thr Gly		
	420	425 430
Val Gly Arg Gly Arg Tyr Met Ser Pro Val Val Val Gly Ser Leu Asn		
	435	440 445
Glu Gln Leu His Asn Pro Glu Leu Asp Thr Ser Ser Pro Asp Val Pro		
	450	455 460
Thr Arg Arg Arg Arg Leu His Leu Arg Ala Ala Thr Ala Arg Met Lys		
	465	470 475 480
Ala Ala Ala Leu Gly Gln Asp Leu Asp Met His Asp Ala Asp Glu Asp		
	485	490 495
Ala Ser Gly Ser Gly Gly Gly Gln Gln Tyr Ala Asp Asp Trp Lys Ala		
	500	505 510

Gly Ala Ala Pro Val Val Pro Pro Ala Arg Pro Pro Arg Pro Pro Arg
 515 520 525

Pro Pro Arg Arg Asp Gly Leu Gly Val Arg Gly Gly Ser Gly Ser Ala
 530 535 540

Arg Tyr Asn Gln Gly Arg Ser Arg Asn Leu Gly Ser Ser Val Gly Leu
 545 550 555 560

His Ala Pro Arg Val Phe Ile Leu Leu Pro Ser Ala Leu Thr Leu Leu
 565 570 575

Gly Leu Arg

<210> 15
 <211> 579
 <212> PRT
 <213> Homo sapiens

<400> 15

Met Ser Ala Leu Arg Pro Leu Leu Leu Leu Leu Pro Leu Cys Pro
 1 5 10 15

Gly Pro Gly Pro Gly Pro Gly Ser Glu Ala Lys Val Thr Arg Ser Cys
 20 25 30

Ala Glu Thr Arg Gln Val Leu Gly Ala Arg Gly Tyr Ser Leu Asn Leu
 35 40 45

Ile Pro Pro Ala Leu Ile Ser Gly Glu His Leu Arg Val Cys Pro Gln
 50 55 60

Glu Tyr Thr Cys Cys Ser Ser Glu Thr Glu Gln Arg Leu Ile Arg Glu
 65 70 75 80

Thr Glu Ala Thr Phe Arg Gly Leu Val Glu Asp Ser Gly Ser Phe Leu
 85 90 95

Val His Thr Leu Ala Ala Arg His Arg Lys Phe Asp Glu Phe Phe Leu
 100 105 110

Glu Met Leu Ser Val Ala Gln His Ser Leu Thr Gln Leu Phe Ser His
 115 120 125

Ser Tyr Gly Arg Leu Tyr Ala Gln His Ala Leu Ile Phe Asn Gly Leu
 130 135 140

Phe Ser Arg Leu Arg Asp Phe Tyr Gly Glu Ser Gly Glu Gly Leu Asp
145 150 155 160

Asp Thr Leu Ala Asp Phe Trp Ala Gln Leu Leu Glu Arg Val Phe Pro
165 170 175

Leu Leu His Pro Gln Tyr Ser Phe Pro Pro Asp Tyr Leu Leu Cys Leu
180 185 190

Ser Arg Leu Ala Ser Ser Thr Asp Gly Ser Leu Gln Pro Phe Gly Asp
195 200 205

Ser Pro Arg Arg Leu Arg Leu Gln Ile Thr Arg Thr Leu Val Ala Ala
210 215 220

Arg Ala Phe Val Gln Gly Leu Glu Thr Gly Arg Asn Val Val Ser Glu
225 230 235 240

Ala Leu Lys Val Pro Val Ser Glu Gly Cys Ser Gln Ala Leu Met Arg
245 250 255

Leu Ile Gly Cys Pro Leu Cys Arg Gly Val Pro Ser Leu Met Pro Cys
260 265 270

Gln Gly Phe Cys Leu Asn Val Val Arg Gly Cys Leu Ser Ser Arg Gly
275 280 285

Leu Glu Pro Asp Trp Gly Asn Tyr Leu Asp Gly Leu Leu Ile Leu Ala
290 295 300

Asp Lys Leu Gln Gly Pro Phe Ser Phe Glu Leu Thr Ala Glu Ser Ile
305 310 315 320

Gly Val Lys Ile Ser Glu Gly Leu Met Tyr Leu Gln Glu Asn Ser Ala
325 330 335

Lys Val Ser Ala Gln Val Phe Gln Glu Cys Gly Pro Pro Asp Pro Val
340 345 350

Pro Ala Arg Asn Arg Arg Ala Pro Pro Pro Arg Glu Glu Ala Gly Arg
355 360 365

Leu Trp Ser Met Val Thr Glu Glu Glu Arg Pro Thr Thr Ala Ala Gly
370 375 380

Thr Asn Leu His Arg Leu Val Trp Glu Leu Arg Glu Arg Leu Ala Arg
385 390 395 400

Met Arg Gly Phe Trp Ala Arg Leu Ser Leu Thr Val Cys Gly Asp Ser
405 410 415

Arg Met Ala Ala Asp Ala Ser Leu Glu Ala Ala Pro Cys Trp Thr Gly
420 425 430

Ala Gly Arg Gly Arg Tyr Leu Pro Pro Val Val Gly Gly Ser Pro Ala
435 440 445

Glu Gln Val Asn Asn Pro Glu Leu Lys Val Asp Ala Ser Gly Pro Asp
450 455 460

Val Pro Thr Arg Arg Arg Arg Leu Gln Leu Arg Ala Ala Thr Ala Arg
465 470 475 480

Met Lys Thr Ala Ala Leu Gly His Asp Leu Asp Gly Gln Asp Ala Asp
485 490 495

Glu Asp Ala Ser Gly Ser Gly Gly Gly Gln Gln Tyr Ala Asp Asp Trp
500 505 510

Met Ala Gly Ala Val Ala Pro Pro Ala Arg Pro Pro Arg Pro Pro Tyr
515 520 525

Pro Pro Arg Arg Asp Gly Ser Gly Gly Lys Gly Gly Gly Gly Ser Ala
530 535 540

Arg Tyr Asn Gln Gly Arg Ser Arg Ser Gly Gly Ala Ser Ile Gly Phe
545 550 555 560

His Thr Gln Thr Ile Leu Ile Leu Ser Leu Ser Ala Leu Ala Leu Leu
565 570 575

Gly Pro Arg

<210> 16
<211> 580
<212> PRT
<213> Homo sapiens

<400> 16

Met Ala Gly Thr Val Arg Thr Ala Cys Leu Val Val Ala Met Leu Leu
1 5 10 15

Ser Leu Asp Phe Pro Gly Gln Ala Gln Pro Pro Pro Pro Pro Pro Asp
20 25 30

Ala Thr Cys His Gln Val Arg Ser Phe Phe Gln Arg Leu Gln Pro Gly
35 40 45

Leu Lys Trp Val Pro Glu Thr Pro Val Pro Gly Ser Asp Leu Gln Val
50 55 60

Cys Leu Pro Lys Gly Pro Thr Cys Cys Ser Arg Lys Met Glu Glu Lys
65 70 75 80

Tyr Gln Leu Thr Ala Arg Leu Asn Met Glu Gln Leu Leu Gln Ser Ala
85 90 95

Ser Met Glu Leu Lys Phe Leu Ile Ile Gln Asn Ala Ala Val Phe Gln
100 105 110

Glu Ala Phe Glu Ile Val Val Arg His Ala Lys Asn Tyr Thr Asn Ala
115 120 125

Met Phe Lys Asn Asn Tyr Pro Ser Leu Thr Pro Gln Ala Phe Glu Phe
130 135 140

Val Gly Glu Phe Phe Thr Asp Val Ser Leu Tyr Ile Leu Gly Ser Asp
145 150 155 160

Ile Asn Val Asp Asp Met Val Asn Glu Leu Phe Asp Ser Leu Phe Pro
165 170 175

Val Ile Tyr Thr Gln Leu Met Asn Pro Gly Leu Pro Asp Ser Ala Leu
180 185 190

Asp Ile Asn Glu Cys Leu Arg Gly Ala Arg Arg Asp Leu Lys Val Phe
195 200 205

Gly Asn Phe Pro Lys Leu Ile Met Thr Gln Val Ser Lys Ser Leu Gln
210 215 220

Val Thr Arg Ile Phe Leu Gln Ala Leu Asn Leu Gly Ile Glu Val Ile
225 230 235 240

Asn Thr Thr Asp His Leu Lys Phe Ser Lys Asp Cys Gly Arg Met Leu
245 250 255

Thr Arg Met Trp Tyr Cys Ser Tyr Cys Gln Gly Leu Met Met Val Lys

260	265	270
Pro Cys Gly Gly Tyr Cys Asn Val Val Met Gln Gly Cys Met Ala Gly		
275	280	285
Val Val Glu Ile Asp Lys Tyr Trp Arg Glu Tyr Ile Leu Ser Leu Glu		
290	295	300
<hr/>		
Glu Leu Val Asn Gly Met Tyr Arg Ile Tyr Asp Met Glu Asn Val Leu		
305	310	315
Leu Gly Leu Phe Ser Thr Ile His Asp Ser Ile Gln Tyr Val Gln Lys		
	325	330
Asn Ala Gly Lys Leu Thr Thr Thr Ile Gly Lys Leu Cys Ala His Ser		
	340	345
Gln Gln Arg Gln Tyr Arg Ser Ala Tyr Tyr Pro Glu Asp Leu Phe Ile		
	355	360
Asp Lys Lys Val Leu Lys Val Ala His Val Glu His Glu Glu Thr Leu		
	370	375
Ser Ser Arg Arg Arg Glu Leu Ile Gln Lys Leu Lys Ser Phe Ile Ser		
385	390	395
Phe Tyr Ser Ala Leu Pro Gly Tyr Ile Cys Ser His Ser Pro Val Ala		
	405	410
Glu Asn Asp Thr Leu Cys Trp Asn Gly Gln Glu Leu Val Glu Arg Tyr		
	420	425
Ser Gln Lys Ala Ala Arg Asn Gly Met Lys Asn Gln Phe Asn Leu His		
	435	440
Glu Leu Lys Met Lys Gly Pro Glu Pro Val Val Ser Gln Ile Ile Asp		
450	455	460
Lys Leu Lys His Ile Asn Gln Leu Leu Arg Thr Met Ser Met Pro Lys		
465	470	475
Gly Arg Val Leu Asp Lys Asn Leu Asp Glu Glu Gly Phe Glu Ser Gly		
	485	490
Asp Cys Gly Asp Asp Glu Asp Glu Cys Ile Gly Gly Ser Gly Asp Gly		
	500	505
		510

Met Ile Lys Val Lys Asn Gln Leu Arg Phe Leu Ala Glu Leu Ala Tyr
 515 520 525

Asp Leu Asp Val Asp Asp Ala Pro Gly Asn Ser Gln Gln Ala Thr Pro
 530 535 540

~~Lys Asp Asn Glu Ile Ser Thr Phe His Asn Leu Gly Asn Val His Ser~~
 545 550 555 560

Pro Leu Lys Leu Leu Thr Ser Met Ala Ile Ser Val Val Cys Phe Phe
 565 570 575

Phe Leu Val His
 580

<210> 17
 <211> 580
 <212> PRT
 <213> Homo sapiens

<400> 17

Met Ala Gly Thr Val Arg Thr Ala Cys Leu Val Val Ala Met Leu Leu
 1 5 10 15

Ser Leu Asp Phe Pro Gly Gln Ala Gln Pro Pro Pro Pro Pro Asp
 20 25 30

Ala Thr Cys His Gln Val Arg Ser Phe Phe Gln Arg Leu Gln Pro Gly
 35 40 45

Leu Lys Trp Val Pro Glu Thr Pro Val Pro Gly Ser Asp Leu Gln Val
 50 55 60

Cys Leu Pro Lys Gly Pro Thr Cys Cys Ser Arg Lys Met Glu Glu Lys
 65 70 75 80

Tyr Gln Leu Thr Ala Arg Leu Asn Met Glu Gln Leu Leu Gln Ser Ala
 85 90 95

Ser Met Glu Leu Lys Phe Leu Ile Ile Gln Asn Ala Ala Val Phe Gln
 100 105 110

Glu Ala Phe Glu Ile Val Val Arg His Ala Lys Asn Tyr Thr Asn Ala
 115 120 125

Met Phe Lys Asn Asn Tyr Pro Ser Leu Thr Pro Gln Ala Phe Glu Phe
 130 135 140

Val Gly Glu Phe Phe Thr Asp Val Ser Leu Tyr Ile Leu Gly Ser Asp
145 150 155 160

Ile Asn Val Asp Asp Met Val Asn Glu Leu Phe Asp Ser Leu Phe Pro
165 170 175

Val Ile Tyr Thr Gln Leu Met Asn Pro Gly Leu Pro Asp Ser Ala Leu
180 185 190

Asp Ile Asn Glu Cys Leu Arg Gly Ala Arg Arg Asp Leu Lys Val Phe
195 200 205

Gly Asn Phe Pro Lys Leu Ile Met Thr Gln Val Ser Lys Ser Leu Gln
210 215 220

Val Thr Arg Ile Phe Leu Gln Ala Leu Asn Leu Gly Ile Glu Val Ile
225 230 235 240

Asn Thr Thr Asp His Leu Lys Phe Ser Lys Asp Cys Gly Arg Met Leu
245 250 255

Thr Arg Met Trp Tyr Cys Ser Tyr Cys Gln Gly Leu Met Met Val Lys
260 265 270

Pro Cys Gly Gly Tyr Cys Asn Val Val Met Gln Gly Cys Met Ala Gly
275 280 285

Val Val Glu Ile Asp Lys Tyr Trp Arg Glu Tyr Ile Leu Ser Leu Glu
290 295 300

Glu Leu Val Asn Gly Met Tyr Arg Ile Tyr Asp Met Glu Asn Val Leu
305 310 315 320

Leu Gly Leu Phe Ser Thr Ile His Asp Ser Ile Gln Tyr Val Gln Lys
325 330 335

Asn Ala Gly Lys Leu Thr Thr Thr Ile Gly Lys Leu Cys Ala His Ser
340 345 350

Gln Gln Arg Gln Tyr Arg Ser Ala Tyr Tyr Pro Glu Asp Leu Phe Ile
355 360 365

Asp Lys Lys Val Leu Lys Val Ala His Val Glu His Glu Glu Thr Leu
370 375 380

Ser Ser Arg Arg Arg Glu Leu Ile Gln Lys Leu Lys Ser Phe Ile Ser
385 390 395 400

Phe Tyr Ser Ala Leu Pro Gly Tyr Ile Cys Ser His Ser Pro Val Ala
405 410 415

Glu Asn Asp Thr Leu Cys Trp Asn Gly Gln Glu Leu Val Glu Arg Tyr
420 425 430

Ser Gln Lys Ala Ala Arg Asn Gly Met Lys Asn Gln Phe Asn Leu His
435 440 445

Glu Leu Lys Met Lys Gly Pro Glu Pro Val Val Ser Gln Ile Ile Asp
450 455 460

Lys Leu Lys His Ile Asn Gln Leu Leu Arg Thr Met Ser Met Pro Lys
465 470 475 480

Gly Arg Val Leu Asp Lys Asn Leu Asp Glu Glu Gly Phe Glu Ser Gly
485 490 495

Asp Cys Gly Asp Asp Glu Asp Glu Cys Ile Gly Gly Ser Gly Asp Gly
500 505 510

Met Ile Lys Val Lys Asn Gln Leu Arg Phe Leu Ala Glu Leu Ala Tyr
515 520 525

Asp Leu Asp Val Asp Asp Ala Pro Gly Asn Ser Gln Gln Ala Thr Pro
530 535 540

Lys Asp Asn Glu Ile Ser Thr Phe His Asn Leu Gly Asn Val His Ser
545 550 555 560

Pro Leu Lys Leu Leu Thr Ser Met Ala Ile Ser Val Val Cys Phe Phe
565 570 575

Phe Leu Val His
580

<210> 18
<211> 556
<212> PRT
<213> Homo sapiens

<400> 18

Met Ala Arg Phe Gly Leu Pro Ala Leu Leu Cys Thr Leu Ala Val Leu
1 5 10 15

Ser Ala Ala Leu Leu Ala Ala Glu Leu Lys Ser Lys Ser Cys Ser Glu
20 25 30

Val Arg Arg Leu Tyr Val Ser Lys Gly Phe Asn Lys Asn Asp Ala Pro
35 40 45

Leu His Glu Ile Asn Gly Asp His Leu Lys Ile Cys Pro Gln Gly Ser
50 55 60

Thr Cys Cys Ser Gln Glu Met Glu Glu Lys Tyr Ser Leu Gln Ser Lys
65 70 75 80

Asp Asp Phe Lys Ser Val Val Ser Glu Gln Cys Asn His Leu Gln Ala
85 90 95

Val Phe Ala Ser Arg Tyr Lys Lys Phe Asp Glu Phe Phe Lys Glu Leu
100 105 110

Leu Glu Asn Ala Glu Lys Ser Leu Asn Asp Met Phe Val Lys Thr Tyr
115 120 125

Gly His Leu Tyr Met Gln Asn Ser Glu Leu Phe Lys Asp Leu Phe Val
130 135 140

Glu Leu Lys Arg Tyr Tyr Val Val Gly Asn Val Asn Leu Glu Glu Met
145 150 155 160

Leu Asn Asp Phe Trp Ala Arg Leu Leu Glu Arg Met Phe Arg Leu Val
165 170 175

Asn Ser Gln Tyr His Phe Thr Asp Glu Tyr Leu Glu Cys Val Ser Lys
180 185 190

Tyr Thr Glu Gln Leu Lys Pro Phe Gly Asp Val Pro Arg Lys Leu Lys
195 200 205

Leu Gln Val Thr Arg Ala Phe Val Ala Ala Arg Thr Phe Ala Gln Gly
210 215 220

Leu Ala Val Ala Gly Asp Val Val Ser Lys Val Ser Val Val Asn Pro
225 230 235 240

Thr Ala Gln Cys Thr His Ala Leu Leu Lys Met Ile Tyr Cys Ser His
245 250 255

Cys Arg Gly Leu Val Thr Val Lys Pro Cys Tyr Asn Tyr Cys Ser Asn

260	265	270
Ile Met Arg Gly Cys Leu Ala Asn Gln Gly Asp Leu Asp Phe Glu Trp		
275	280	285
Asn Asn Phe Ile Asp Ala Met Leu Met Val Ala Glu Arg Leu Glu Gly		
290	295	300
<hr/>		
Pro Phe Asn Ile Glu Ser Val Met Asp Pro Ile Asp Val Lys Ile Ser		
305	310	315
Asp Ala Ile Met Asn Met Gln Asp Asn Ser Val Gln Val Ser Gln Lys		
325	330	335
Val Phe Gln Gly Cys Gly Pro Pro Lys Pro Leu Pro Ala Gly Arg Ile		
340	345	350
Ser Arg Ser Ile Ser Glu Ser Ala Phe Ser Ala Arg Phe Arg Pro His		
355	360	365
His Pro Glu Glu Arg Pro Thr Thr Ala Ala Gly Thr Ser Leu Asp Arg		
370	375	380
Leu Val Thr Asp Val Lys Glu Lys Leu Lys Gln Ala Lys Lys Phe Trp		
385	390	395
Ser Ser Leu Pro Ser Asn Val Cys Asn Asp Glu Arg Met Ala Ala Gly		
405	410	415
Asn Gly Asn Glu Asp Asp Cys Trp Asn Gly Lys Gly Lys Ser Arg Tyr		
420	425	430
Leu Phe Ala Val Thr Gly Asn Gly Leu Ala Asn Gln Gly Asn Asn Pro		
435	440	445
Glu Val Gln Val Asp Thr Ser Lys Pro Asp Ile Leu Ile Leu Arg Gln		
450	455	460
Ile Met Ala Leu Arg Val Met Thr Ser Lys Met Lys Asn Ala Tyr Asn		
465	470	475
Gly Asn Asp Val Asp Phe Phe Asp Ile Ser Asp Glu Ser Ser Gly Glu		
485	490	495
Gly Ser Gly Ser Gly Cys Glu Tyr Gln Gln Cys Pro Ser Glu Phe Asp		
500	505	510

Tyr Asn Ala Thr Asp His Ala Gly Lys Ser Ala Asn Glu Lys Ala Asp
 515 520 525

Ser Ala Gly Val Arg Pro Gly Ala Gln Ala Tyr Leu Leu Thr Val Phe
 530 535 540

Cys Ile Leu Phe Leu Val Met Gln Arg Glu Trp Arg
 545 550 555

<210> 19
 <211> 556
 <212> PRT
 <213> Homo sapiens

<400> 19

Met Ala Arg Phe Gly Leu Pro Ala Leu Leu Cys Thr Leu Ala Val Leu
 1 5 10 15

Ser Ala Ala Leu Leu Ala Ala Glu Leu Lys Ser Lys Ser Cys Ser Glu
 20 25 30

Val Arg Arg Leu Tyr Val Ser Lys Gly Phe Asn Lys Asn Asp Ala Pro
 35 40 45

Leu His Glu Ile Asn Gly Asp His Leu Lys Ile Cys Pro Gln Gly Ser
 50 55 60

Thr Cys Cys Ser Gln Glu Met Glu Glu Lys Tyr Ser Leu Gln Ser Lys
 65 70 75 80

Asp Asp Phe Lys Ser Val Val Ser Glu Gln Cys Asn His Leu Gln Ala
 85 90 95

Val Phe Ala Ser Arg Tyr Lys Lys Phe Asp Glu Phe Phe Lys Glu Leu
 100 105 110

Leu Glu Asn Ala Glu Lys Ser Leu Asn Asp Met Phe Val Lys Thr Tyr
 115 120 125

Gly His Leu Tyr Met Gln Asn Ser Glu Leu Phe Lys Asp Leu Phe Val
 130 135 140

Glu Leu Lys Arg Tyr Tyr Val Val Gly Asn Val Asn Leu Glu Glu Met
 145 150 155 160

Leu Asn Asp Phe Trp Ala Arg Leu Leu Glu Arg Met Phe Arg Leu Val
 165 170 175

Asn Ser Gln Tyr His Phe Thr Asp Glu Tyr Leu Glu Cys Val Ser Lys
 180 185 190

Tyr Thr Glu Gln Leu Lys Pro Phe Gly Asp Val Pro Arg Lys Leu Lys
 195 200 205

Leu Gln Val Thr Arg Ala Phe Val Ala Ala Arg Thr Phe Ala Gln Gly
 210 215 220

Leu Ala Val Ala Gly Asp Val Val Ser Lys Val Ser Val Val Asn Pro
 225 230 235 240

Thr Ala Gln Cys Thr His Ala Leu Leu Lys Met Ile Tyr Cys Ser His
 245 250 255

Cys Arg Gly Leu Val Thr Val Lys Pro Cys Tyr Asn Tyr Cys Ser Asn
 260 265 270

Ile Met Arg Gly Cys Leu Ala Asn Gln Gly Asp Leu Asp Phe Glu Trp
 275 280 285

Asn Asn Phe Ile Asp Ala Met Leu Met Val Ala Glu Arg Leu Glu Gly
 290 295 300

Pro Phe Asn Ile Glu Ser Val Met Asp Pro Ile Asp Val Lys Ile Ser
 305 310 315 320

Asp Ala Ile Met Asn Met Gln Asp Asn Ser Val Gln Val Ser Gln Lys
 325 330 335

Val Phe Gln Gly Cys Gly Pro Pro Lys Pro Leu Pro Ala Gly Arg Ile
 340 345 350

Ser Arg Ser Ile Ser Glu Ser Ala Phe Ser Ala Arg Phe Arg Pro His
 355 360 365

His Pro Glu Glu Arg Pro Thr Thr Ala Ala Gly Thr Ser Leu Asp Arg
 370 375 380

Leu Val Thr Asp Val Lys Asp Lys Leu Lys Gln Ala Lys Lys Phe Trp
 385 390 395 400

Ser Ser Leu Pro Ser Asn Val Cys Asn Asp Glu Arg Met Ala Ala Gly
 405 410 415

Asn Gly Asn Glu Asp Asp Cys Trp Asn Gly Lys Gly Lys Ser Arg Tyr
 420 425 430

Leu Phe Ala Val Thr Gly Asn Gly Leu Val Asn Gln Gly Asn Asn Pro
 435 440 445

Glu Val Gln Val Asp Thr Ser Lys Pro Asp Ile Leu Ile Leu Arg Gln
 450 455 460

Ile Met Ala Leu Arg Val Met Thr Ser Lys Met Lys Asn Ala Tyr Asn
 465 470 475 480

Gly Asn Asp Val Asp Phe Phe Asp Ile Ser Asp Glu Ser Ser Gly Glu
 485 490 495

Gly Ser Gly Ser Gly Cys Glu Tyr Gln Gln Cys Pro Ser Glu Phe Asp
 500 505 510

Tyr Asn Ala Thr Asp His Ala Gly Lys Ser Ala Asn Glu Lys Ala Asp
 515 520 525

Ser Ala Gly Val Arg Pro Gly Ala Gln Ala Tyr Leu Leu Thr Val Phe
 530 535 540

Cys Ile Leu Phe Leu Val Met Gln Arg Glu Trp Arg
 545 550 555

<210> 20
 <211> 572
 <212> PRT
 <213> Homo sapiens

<400> 20

Met Asp Ala Gln Thr Trp Pro Val Gly Phe Arg Cys Leu Leu Leu Leu
 1 5 10 15

Ala Leu Val Gly Ser Ala Arg Ser Glu Gly Val Gln Thr Cys Glu Glu
 20 25 30

Val Arg Lys Leu Phe Gln Trp Arg Leu Leu Gly Ala Val Arg Gly Leu
 35 40 45

Pro Asp Ser Pro Arg Ala Gly Pro Asp Leu Gln Val Cys Ile Ser Lys
 50 55 60

Lys Pro Thr Cys Cys Thr Arg Lys Met Glu Glu Arg Tyr Gln Ile Ala
 65 70 75 80

Ala Arg Gln Asp Met Gln Gln Phe Leu Gln Thr Ser Ser Ser Thr Leu
85 90 95

Lys Phe Leu Ile Ser Arg Asn Ala Ala Ala Phe Gln Glu Thr Leu Glu
100 105 110

Thr Leu Ile Lys Gln Ala Glu Asn Tyr Thr Ser Ile Leu Phe Cys Ser
115 120 125

Thr Tyr Arg Asn Met Ala Leu Glu Ala Ala Ala Ser Val Gln Glu Phe
130 135 140

Phe Thr Asp Val Gly Leu Tyr Leu Phe Gly Ala Asp Val Asn Pro Glu
145 150 155 160

Glu Phe Val Asn Arg Phe Phe Asp Ser Leu Phe Pro Leu Val Tyr Asn
165 170 175

His Leu Ile Asn Pro Gly Val Thr Asp Ser Ser Leu Glu Tyr Ser Glu
180 185 190

Cys Ile Arg Met Ala Arg Arg Asp Val Ser Pro Phe Gly Asn Ile Pro
195 200 205

Gln Arg Val Met Gly Gln Met Gly Arg Ser Leu Leu Pro Ser Arg Thr
210 215 220

Phe Leu Gln Ala Leu Asn Leu Gly Ile Glu Val Ile Asn Thr Thr Asp
225 230 235 240

Tyr Leu His Phe Ser Lys Glu Cys Ser Arg Ala Leu Leu Lys Met Gln
245 250 255

Tyr Cys Pro His Cys Gln Gly Leu Ala Leu Thr Lys Pro Cys Met Gly
260 265 270

Tyr Cys Leu Asn Val Met Arg Gly Cys Leu Ala His Met Ala Glu Leu
275 280 285

Asn Pro His Trp His Ala Tyr Ile Arg Ser Leu Glu Glu Leu Ser Asp
290 295 300

Ala Met His Gly Thr Tyr Asp Ile Gly His Val Leu Leu Asn Phe His
305 310 315 320

Leu Leu Val Asn Asp Ala Val Leu Gln Ala His Leu Asn Gly Gln Lys

	325		330		335										
Leu	Leu	Glu	Gln	Val	Asn	Arg	Ile	Cys	Gly	Arg	Pro	Val	Arg	Thr	Pro
	340							345					350		
Thr	Gln	Ser	Pro	Arg	Cys	Ser	Phe	Asp	Gln	Ser	Lys	Glu	Lys	His	Gly
	355						360					365			
Met	Lys	Thr	Thr	Thr	Arg	Asn	Ser	Glu	Glu	Thr	Leu	Ala	Asn	Arg	Arg
	370						375				380				
Lys	Glu	Phe	Ile	Asn	Ser	Leu	Arg	Leu	Tyr	Arg	Ser	Phe	Tyr	Gly	Gly
385					390					395					400
Leu	Ala	Asp	Gln	Leu	Cys	Ala	Asn	Glu	Leu	Ala	Ala	Ala	Asp	Gly	Leu
				405					410					415	
Pro	Cys	Trp	Asn	Gly	Glu	Asp	Ile	Val	Lys	Ser	Tyr	Thr	Gln	Arg	Val
			420					425					430		
Val	Gly	Asn	Gly	Ile	Lys	Ala	Gln	Ser	Gly	Asn	Pro	Glu	Val	Lys	Val
		435					440					445			
Lys	Gly	Ile	Asp	Pro	Val	Ile	Asn	Gln	Ile	Ile	Asp	Lys	Leu	Lys	His
	450						455				460				
Val	Val	Gln	Leu	Leu	Gln	Gly	Arg	Ser	Pro	Lys	Pro	Asp	Lys	Trp	Glu
465					470					475					480
Leu	Leu	Gln	Leu	Gly	Ser	Gly	Gly	Gly	Met	Val	Glu	Gln	Val	Ser	Gly
				485					490					495	
Asp	Cys	Asp	Asp	Glu	Asp	Gly	Cys	Gly	Gly	Ser	Gly	Ser	Gly	Glu	Val
			500					505						510	
Lys	Arg	Thr	Leu	Lys	Ile	Thr	Asp	Trp	Met	Pro	Asp	Asp	Met	Asn	Phe
		515					520					525			
Ser	Asp	Val	Lys	Gln	Ile	His	Gln	Thr	Asp	Thr	Gly	Ser	Thr	Leu	Asp
	530					535					540				
Thr	Thr	Gly	Ala	Gly	Cys	Ala	Val	Ala	Thr	Glu	Ser	Met	Thr	Phe	Thr
545					550					555					560
Leu	Ile	Ser	Val	Val	Met	Leu	Leu	Pro	Gly	Ile	Trp				
				565					570						

<210> 21
 <211> 555
 <212> PRT
 <213> Homo sapiens

<400> 21

Met Pro Ser Trp Ile Gly Ala Val Ile Leu Pro Leu Leu Gly Leu Leu
 1 5 10 15

Leu Ser Leu Pro Ala Gly Ala Asp Val Lys Ala Arg Ser Cys Gly Glu
 20 25 30

Val Arg Gln Ala Tyr Gly Ala Lys Gly Phe Ser Leu Ala Asp Ile Pro
 35 40 45

Tyr Gln Glu Ile Ala Gly Glu His Leu Arg Ile Cys Pro Gln Glu Tyr
 50 55 60

Thr Cys Cys Thr Thr Glu Met Glu Asp Lys Leu Ser Gln Gln Ser Lys
 65 70 75 80

Leu Glu Phe Glu Asn Leu Val Glu Glu Thr Ser His Phe Val Arg Thr
 85 90 95

Thr Phe Val Ser Arg His Lys Lys Phe Asp Glu Phe Phe Arg Glu Leu
 100 105 110

Leu Glu Asn Ala Glu Lys Ser Leu Asn Asp Met Phe Val Arg Thr Tyr
 115 120 125

Gly Met Leu Tyr Met Gln Asn Ser Glu Val Phe Gln Asp Leu Phe Thr
 130 135 140

Glu Leu Lys Arg Tyr Tyr Thr Gly Gly Asn Val Asn Leu Glu Glu Met
 145 150 155 160

Leu Asn Asp Phe Trp Ala Arg Leu Leu Glu Arg Met Phe Gln Leu Ile
 165 170 175

Asn Pro Gln Tyr His Phe Ser Glu Asp Tyr Leu Glu Cys Val Ser Lys
 180 185 190

Tyr Thr Asp Gln Leu Lys Pro Phe Gly Asp Val Pro Arg Lys Leu Lys
 195 200 205

Ile Gln Val Thr Arg Ala Phe Ile Ala Ala Arg Thr Phe Val Gln Gly
 210 215 220

Leu Thr Val Gly Arg Glu Val Ala Asn Arg Val Ser Lys Val Ser Pro
 225 230 235 240

Thr Pro Gly Cys Ile Arg Ala Leu Met Lys Met Leu Tyr Cys Pro Tyr
 245 250 255

Cys Arg Gly Leu Pro Thr Val Arg Pro Cys Asn Asn Tyr Cys Leu Asn
 260 265 270

Val Met Lys Gly Cys Leu Ala Asn Gln Ala Asp Leu Asp Thr Glu Trp
 275 280 285

Asn Leu Phe Ile Asp Ala Met Leu Leu Val Ala Glu Arg Leu Glu Gly
 290 295 300

Pro Phe Asn Ile Glu Ser Val Met Asp Pro Ile Asp Val Lys Ile Ser
 305 310 315 320

Glu Ala Ile Met Asn Met Gln Glu Asn Ser Met Gln Val Ser Ala Lys
 325 330 335

Val Phe Gln Gly Cys Gly Gln Pro Lys Pro Ala Pro Ala Leu Arg Ser
 340 345 350

Ala Arg Ser Ala Pro Glu Asn Phe Asn Thr Arg Phe Arg Pro Tyr Asn
 355 360 365

Pro Glu Glu Arg Pro Thr Thr Ala Ala Gly Thr Ser Leu Asp Arg Leu
 370 375 380

Val Thr Asp Ile Lys Glu Lys Leu Lys Leu Ser Lys Lys Val Trp Ser
 385 390 395 400

Ala Leu Pro Tyr Thr Ile Cys Lys Asp Glu Ser Val Thr Ala Gly Thr
 405 410 415

Ser Asn Glu Glu Glu Cys Trp Asn Gly His Ser Lys Ala Arg Tyr Leu
 420 425 430

Pro Glu Ile Met Asn Asp Gly Leu Thr Asn Gln Ile Asn Asn Pro Glu
 435 440 445

Val Asp Val Asp Ile Thr Arg Pro Asp Thr Phe Ile Arg Gln Gln Ile
 450 455 460

Met Ala Leu Arg Val Met Thr Asn Lys Leu Lys Asn Ala Tyr Asn Gly
 465 470 475 480

Asn Asp Val Asn Phe Gln Asp Thr Ser Asp Glu Ser Ser Gly Ser Gly
 485 490 495

Ser Gly Ser Gly Cys Met Asp Asp Val Cys Pro Thr Glu Phe Glu Phe
 500 505 510

Val Thr Thr Glu Ala Pro Ala Val Asp Pro Asp Arg Arg Glu Val Asp
 515 520 525

Ser Ser Ala Ala Gln Arg Gly His Ser Leu Leu Ser Trp Ser Leu Thr
 530 535 540

Cys Ile Val Leu Ala Leu Gln Arg Leu Cys Arg
 545 550 555

<210> 22
 <211> 555
 <212> PRT
 <213> Homo sapiens

<400> 22

Met Pro Ser Trp Ile Gly Ala Val Ile Leu Pro Leu Leu Gly Leu Leu
 1 5 10 15

Leu Ser Leu Pro Ala Gly Ala Asp Val Lys Ala Arg Ser Cys Gly Glu
 20 25 30

Val Arg Gln Ala Tyr Gly Ala Lys Gly Phe Ser Leu Ala Asp Ile Pro
 35 40 45

Tyr Gln Glu Ile Ala Gly Glu His Leu Arg Ile Cys Pro Gln Glu Tyr
 50 55 60

Thr Cys Cys Thr Thr Glu Met Glu Asp Lys Leu Ser Gln Gln Ser Lys
 65 70 75 80

Leu Glu Phe Glu Asn Leu Val Glu Glu Thr Ser His Phe Val Arg Thr
 85 90 95

Thr Phe Val Ser Arg His Lys Lys Phe Asp Glu Phe Phe Arg Glu Leu
 100 105 110

Leu Glu Asn Ala Glu Lys Ser Leu Asn Asp Met Phe Val Arg Thr Tyr
 115 120 125

Gly Met Leu Tyr Met Gln Asn Ser Glu Val Phe Gln Asp Leu Phe Thr
 130 135 140

Glu Leu Lys Arg Tyr Tyr Thr Gly Gly Asn Val Asn Leu Glu Glu Met
 145 150 155 160

~~Leu Asn Asp Phe Trp Ala Arg Leu Leu Glu Arg Met Phe Gln Leu Ile~~
 165 170 175

Asn Pro Gln Tyr His Phe Ser Glu Asp Tyr Leu Glu Cys Val Ser Lys
 180 185 190

Tyr Thr Asp Gln Leu Lys Pro Phe Gly Asp Val Pro Arg Lys Leu Lys
 195 200 205

Ile Gln Val Thr Arg Ala Phe Ile Ala Ala Arg Thr Phe Val Gln Gly
 210 215 220

Leu Thr Val Gly Arg Glu Val Ala Asn Arg Val Ser Lys Val Ser Pro
 225 230 235 240

Thr Pro Gly Cys Ile Arg Ala Leu Met Lys Met Leu Tyr Cys Pro Tyr
 245 250 255

Cys Arg Gly Leu Pro Thr Val Arg Pro Cys Asn Asn Tyr Cys Leu Asn
 260 265 270

Val Met Lys Gly Cys Leu Ala Asn Gln Ala Asp Leu Asp Thr Glu Trp
 275 280 285

Asn Leu Phe Ile Asp Ala Met Leu Leu Val Ala Glu Arg Leu Glu Gly
 290 295 300

Pro Phe Asn Ile Glu Ser Val Met Asp Pro Ile Asp Val Lys Ile Ser
 305 310 315 320

Glu Ala Ile Met Asn Met Gln Glu Asn Ser Met Gln Val Ser Ala Lys
 325 330 335

Val Phe Gln Gly Cys Gly Gln Pro Lys Pro Ala Pro Ala Leu Arg Ser
 340 345 350

Ala Arg Ser Ala Pro Glu Asn Phe Asn Thr Arg Phe Arg Pro Tyr Asn
 355 360 365

Pro Glu Glu Arg Pro Thr Thr Ala Ala Gly Thr Ser Leu Asp Arg Leu

370 375 380
 Val Thr Asp Ile Lys Glu Lys Leu Lys Leu Ser Lys Lys Val Trp Ser
 385 390 395 400
 Ala Leu Pro Tyr Thr Ile Cys Lys Asp Glu Ser Val Thr Ala Gly Thr
 405 410 415
 Ser Asn Glu Glu Glu Cys Trp Asn Gly His Ser Lys Ala Arg Tyr Leu
 420 425 430
 Pro Glu Ile Met Asn Asp Gly Leu Thr Asn Gln Ile Asn Asn Pro Glu
 435 440 445
 Val Asp Val Asp Ile Thr Arg Pro Asp Thr Phe Ile Arg Gln Gln Ile
 450 455 460
 Met Ala Leu Arg Val Met Thr Asn Lys Leu Lys Asn Ala Tyr Asn Gly
 465 470 475 480
 Asn Asp Val Asn Phe Gln Asp Thr Ser Asp Glu Ser Ser Gly Ser Gly
 485 490 495
 Ser Gly Ser Gly Cys Met Asp Asp Val Cys Pro Thr Glu Phe Glu Phe
 500 505 510
 Val Thr Thr Glu Ala Pro Ala Val Asp Pro Asp Arg Arg Glu Val Asp
 515 520 525
 Ser Ser Ala Ala Gln Arg Gly His Ser Leu Leu Ser Trp Ser Leu Thr
 530 535 540
 Cys Ile Val Leu Ala Leu Gln Arg Leu Cys Arg
 545 550 555

<210> 23
 <211> 450
 <212> DNA
 <213> Homo sapiens

<400> 23
 ttttttttctt tcttttcttc cttcttttctt tgtttctttc tctgtctctc tctctctctt 60
 tctttctctc tctctctttc tttcttgaga tgacgtctag ctatgttgcc ccagctgggc 120
 tccaactcct ggcctcaagt gatcctcctg cctcagcctc tcaaagtgct gtgattacag 180
 gtgtgagcca ccatgcctgg ctccaggctt tcttgaatcc cctcaccag atagttgccc 240
 cagtcaggct ccagtcctct gctgctgaga cagccacgaa ccacgttgag gcagaagccc 300

tggcagggca taagtgaggg gaccccccg caccaggggac agccgatgag acgcatcaga 360
 gcctggctgc agccttcaga caccggcacc tgggggcaga gagtggggct gtgtcctcca 420
 caaacgctgt ggtgctctc agcctggctt 450

<210> 24

<211> 2118

<212> DNA

<213> Homo sapiens

<400> 24

gcgcccaggt agctgcgagg aaacttttgc agcggctggg tagcagcacg tctcttgcctc 60
 ctcagggcca ctgccaggct tgccgagtc tgggactgct ctgcctccgg ctgccactct 120
 ccgcgctct cctagctccc tgcgaagcag gatggccggg accgtgcgca ccgcgtgctt 180
 ggtggtggcg atgctgtca gcttggactt cccgggacag gcgcagcccc cgccgccgcc 240
 gccggacgcc acctgtcacc aagtcgctc cttcttcag agactgcagc ccggactcaa 300
 gtgggtgcca gaaactcccg tgccaggatc agatttgcaa gtatgtctcc ctaagggccc 360
 aacatgctgc tcaagaaaga tggaagaaaa ataccaacta acagcacgat tgaacatgga 420
 acagctgctt cagtctgcaa gtatggagct caagttctta attattcaga atgctgcggt 480
 tttccaagag gcctttgaaa ttgttgttcg ccatgccaaag aactacacca atgccatgtt 540
 caagaacaac taoccaagcc tgactccaca agcttttgag tttgtgggtg aatttttcac 600
 agatgtgtct ctctacatct tgggttctga catcaatgta gatgacatgg tcaatgaatt 660
 gtttgacagc ctgtttccag tcctctatac ccagcnaatg aaccagggcc tgctgattc 720
 agccttggac atcaatgagt gcctccgagg agcaagacgt gacctgaaag tatttgggaa 780
 tttccccaag cttattatga cccaggtttc caagtcactg caagtcacta ggatcttctt 840
 tcaggctctg aatcttggaa ttgaagtgat caacacaact gatcacctga agttcagtaa 900
 ggactgtggc cgaatgtca ccagaatgtg gtactgtctt tactgccagg gactgatgat 960
 ggttaaacc cgtggcggtt actgcaatgt ggtcatgcaa ggctgtatgg cagggtgtgtt 1020
 ggagattgac aagtactgga gagaatacat tctgtccctt gaagaacttg tgaatggcat 1080
 gtacagaatc tatgacatgg agaacgtact gcttggcttc ttttcaacaa tccatgatc 1140
 tatccagtat gtccagaaga atgcaggaaa gctgaccacc actattggca agttatgtgc 1200
 ccattctcaa caacgccaat atagatctgc ttattatcct gaagatctct ttattgacaa 1260
 gaaagtatta aaagtgtctc atgtagaaca tgaagaaacc ttatccagcc gaagaagga 1320
 actaattcag aagttgaagt ctttcatcag cttctatagt gctttgcctg gctacatctg 1380
 cagccatagc cctgtggcgg aaaacgacac cctttgctgg aatggacaag aactcgtgga 1440

gagatacagc	caaaaggcag	caaggaatgg	aatgaaaaac	cagttcaatc	tccatgagct	1500
gaaaatgaag	ggccctgagc	cagtggtcag	tcaaattatt	gacaaactga	agcacattaa	1560
ccagctcctg	agaaccatgt	ctatgcccaa	aggtagagtt	ctggataaaa	acctggatga	1620
ggaagggttt	gaaagtggag	actgcggtga	tgatgaagat	gagtgcattg	gaggctctgg	1680
tgatggaatg	ataaaagtga	agaatcagct	ccgcttcctt	gcagaactgg	cctatgatct	1740
ggatgtggat	gatgcgcctg	gaaacagtca	gcaggcaact	ccgaaggaca	acgagataag	1800
cacctttcac	aacctcggga	acgttcattc	ccgctgaag	cttctcacca	gcattggccat	1860
ctcgggtggtg	tgtttcttct	tcctgggtgca	ctgactgcct	ggtgccccagc	acatgtgctg	1920
ccctacagca	ccctgtggtc	ttcctcgata	aagggaacca	ctttcttatt	tttttctatt	1980
tttttttttt	tgttatcctg	tatacctcct	ccagccatga	agtagaggac	taaccatgtg	2040
ttatgttttc	gaaaatcaaa	tggtatcttt	tggaggaaga	tacattttag	tggtagcata	2100
tagattgtcc	ttttgcaa					2118